THE MATHEMATICAL THEORY OF PROBABILITIES

THE MACMILLAN COMPANY
NEW YORK BOSTON CHICAGO
DALLAS SAN FRANCISCO

MACMILLAN & CO., LIMITED
LONDON BOMBAY CALCUTTA
MELBOURNE

THE MACMILLAN CO. OF CANADA, LTD.
TORONTO

THE MATHEMATICAL THEORY
OF
PROBABILITIES

AND ITS APPLICATION TO
FREQUENCY CURVES AND STATISTICAL
METHODS

BY

ARNE FISHER.

TRANSLATED FROM THE DANISH

BY

CHARLOTTE DICKSON, B.A.
(COLUMBIA)
MATHEMATICAL ASSISTANT IN THE DEPARTMENT OF DEVELOPMENT
AND RESEARCH OF
THE AMERICAN TELEPHONE AND TELEGRAPH COMPANY

AND

WILLIAM BONYNGE, B.A.
(BELFAST)

WITH INTRODUCTORY NOTES

BY

M. C. RORTY

AND

F. W. FRANKLAND, F.I.A., F.A.S, F.S.S.

VOLUME I

MATHEMATICAL PROBABILITIES, FREQUENCY CURVES, HOMOGRADE AND HETEROGRADE STATISTICS

SECOND EDITION GREATLY ENLARGED

NEW YORK
THE MACMILLAN COMPANY
1922

Set up and electrotyped. Published November, 1915.
Second Edition, greatly enlarged, May, 1922.

PRINTED IN THE UNITED STATES OF AMERICA

INTRODUCTORY NOTE TO THE SECOND EDITION.

Mr. Fisher has requested that an introduction be written to this, the second edition of his work on probabilities, which shall indicate some of the practical applications of the mathematical theory with which his treatise deals.

The writer has only a limited knowledge of mathematical technique—yet it has so happened that in twenty-five years of active work as engineer, statistician and executive he has had frequent occasion to call upon the skill of trained mathematicians for the solution of practical problems involving frequency curves and probabilities. Among such mathematicians none has been more helpful, or quicker to perceive the possibility of making valuable applications of higher mathematics to business problems, than Mr. Fisher himself. For this reason it is a duty as well as a privilege to outline, at his request, certain actual practical experiences with mathematical applications and to indicate such possible applications for the future.

The writer's initial experience with frequency curves and probabilities was in the years 1902 and 1903, when it became evident, in analyzing various problems in telephone traffic, that certain peak loads, which were superimposed upon the normal seasonal, weekly, and daily fluctuations, could be accounted for only by the laws of chance. Recourse was, therefore, had to the formulæ then available for approximate summations of the terms of the binomial expansion, and from these a series of curves was drawn which indicated for any given normal hourly traffic (as indicated by studies of seasonal, weekly, and daily variations) the probability that any given short period load would be equalled or exceeded. Practical experience with these curves soon showed that, in spite of minor errors, they were close enough to the real facts to make them of primary importance in traffic studies of all kinds, and particularly in the development of mechanical switching devices. Their use for such purposes has now become a commonplace in telephone engineering.

As a by-product of the preceding application there have been other interesting uses of the same probability curves. Effective studies have been made of the decrease in the total stocks of small machine parts that could be made possible by standardizing and

reducing the number of types of screws, bolts, nuts, etc. The curves can also be applied directly to every line of business and every type of operation where prompt service must be given and where the demand arises from a large number of independent sources, and is, therefore, subject to peak loads determined by the laws of chance, which may be superimposed upon other "normal" peak loads varying with the days of the week, the hours of the day, etc.

Entirely separate applications of frequency curves are those necessary in actuarial work. These are relatively well known. But it is less generally known that one of the most important of business problems, that of depreciation, can be treated effectively only when approached on an actuarial basis with a full understanding of the frequency curves which govern the displacement, year by year, of the physical units involved.

A still further use of frequency curves and the theory of probabilities, which is of immediate practical importance, is in connection with sampling operations. The theory of sampling has already been well developed, but adequate efforts have not yet been made by mathematicians to reduce the processes of sampling to dependable simple rules that can be applied by business executives and statisticians untrained in higher mathematics. In census work, and in statistical and other reports made by business organizations, the waste of money, that could be avoided by an intelligent application of the theory of sampling, is very great. Not only can many reports and analyses be made much more cheaply and quickly by sampling processes, but they can also be made more accurately. Many important items of information can be determined only by trained specialists. In such cases the only procedure, that does not involve prohibitive expense in large census operations, is to tie such items, by a sampling process, to other items which are susceptible of exact enumeration by relatively unskilled enumerators, and then to compute the totals for the special items from the relations of such items to the items which are completely enumerated.

All of the preceding are in the field of immediate practicalities. When we come to the future, one of the most promising uses of mathematics is in the development of logical processes. It is not going too far to say that all business, and most engineering operations are fundamentally based on probabilities. The business man is always dealing in degrees of uncertainty, and even the engineer

has only occasionally a definite set of conditions upon which to base his computations. Where the problem is primarily a financial one, he must balance the cost of overbuilding against the cost of underbuilding; and, if he combines business judgment with engineering skill, he will multiply the amount of each possible loss by the probability of its occurring, and will ordinarily choose, among all possible plans, the plan which involves the minimum probable loss. Here it is not inappropriate to interject the idea that the most practical logic must always be in terms of probabilities, and that a logic which deals, or pretends to deal, in certainties only is not alone useless, but is also harmful and misleading, when difficult problems are to be approached. Such problems can rarely, if ever, be solved except through the cumulation toward a certainty of many small probabilities established from uncorrelated, or only partially correlated, viewpoints.

A final suggestion which is to-day speculative, but may assume important practical aspects in the near future, is with respect to the applications of frequency curves and probabilities to physical and cosmic mathematics. In such mathematics we are forced to assume that all of our measures must arise out of the things measured. When we deal with physical velocities, it would seem that our only measures of velocity can arise out of the velocities themselves. Similar considerations hold true with respect to fundamental measures of physical extension. Under these circumstances we may talk in terms of infinite space and of infinite time, but we can hardly talk in terms of infinites when we are dealing with the dimensions of atomic structure and the velocities of material particles. In these cases it seems very highly probable that we are dealing with frequency distributions which we must measure and define in terms derived from such distributions themselves. With respect to such measures some of our frequency curves may have infinite "tails," but it is more probable that the frequency forms are such that they can be completely defined in finite terms. Along this same line, we may even risk a closing speculation that the relative proportions of organized matter and space in the stellar universe are determined through the operations of the laws of chance in establishing heterogeneities in what is otherwise a homogeneous void-filling medium.

M. C. Rorty.

New York City,
March 2, 1922.

PREFACE TO THE SECOND EDITION.

At the time when the first edition of this little book was published in 1916, I expected to issue a second volume shortly after, dealing with frequency curves and frequency surfaces as well as the related problem of co-variation (correlation). The manuscript for this volume was completed and printing had already commenced on some of the chapters, when a series of misfortunes, not necessarily unexpected, overtook the work. A major part of the manuscript while in transit to a friend in Denmark for review and corrections went down with a Danish vessel when torpedoed by an outlaw German submarine. A duplicate copy was for some reason or other withheld by the British military censor and not returned to the writer until long after the termination of the world war. My third and final copy of the manuscript, which I had submitted to an American friend for critical review was also lost in transit. The veritable nemesis which seems to have followed my efforts is, however, only a verification of the all prevailing laws of chance, which every serious minded student must face with unperturbed attitude. In fact, the above misfortunes have, after all, only made me more determined to complete another collection of notes, which I eventually hope to put into proper shape for publication.

In the meantime the first edition has been out of print for more than two years, and when the publisher asked me to prepare a new edition I took advantage of this opportunity to add several chapters on frequency functions and their application to heterograde statistical series so as to give a complete treatment of statistical functions involving one variable. The book is, therefore, twice its original size and contains the major part of what I originally intended for a second volume.

The reader will readily notice that my treatment of the subject is based throughout upon the principles of the classical probability theory as founded by Bernoulli, De Moivre and above all by the great Laplace and his disciple, Poisson. I am of the opinion that these principles and their further extension by the Scandinavian statisticians and actuaries, Gram, Thiele, Westergaard, Charlier, Wicksell and Jörgensen, offer as yet the best and also the most powerful tools for the treatment of collected statistical data by means of mathematical methods. In the way of adumbration and

economy of thought the Laplacean methods stand unsurpassed in the whole realm of mathematical statistics. I have, therefore, in this volume limited my investigations to a systematic treatment along these lines. I hope, however, in the forthcoming second volume to treat the methods of Pearson, Edgeworth, Kapteyn, Bachelier and Knibbs and show their relation to Laplace's theory.

The reason why the Laplacean doctrine of frequency curves has been ignored until comparatively recent years and has remained more or less obscure is perhaps due to the fact that for more than a century it remained a theory pure and simple and was used but sparingly in practical calculations.

Any statistical theory, in order to be of use in practical work, must be arranged in such a manner that it is readily adaptable to numerical computations. Advanced mathematical computation has not been given its due reward and proper attention in our ordinary academic instruction. A high grade mathematical computer is indeed a "rare bird," much more so in fact than a good mathematician. To arrange and plan the numerical work in connection with the theoretical formulæ so that the detailed and painstaking work is reduced to a minimum, and at the same time afford the proper means for checking and counterchecking, is by no means an easy task and often requires as much ingenuity as the actual development of the theoretical formulæ. While Gauss has always been acknowledged as one of the world's greatest computers and in addition to his extensive work in pure mathematics also did much practical work in surveying, physics, and in financial and actuarial investigations, Laplace during his entire career remained a pure mathematician and apparently failed to grasp the paramount attributes required by a successful computer. His attempt to inject himself into public life, as for instance when he secured for himself an appointment as minister of the interior, must be regarded as a dismal failure as admitted in Napoleon's memorandum on his dismissal.

The failure of Laplace to recognize fully the all-important phase of numerical computations in all observations on statistical mass phenomena is in my opinion the main reason why the Gaussian theory of observations and the allied subject on the theory of least squares has hitherto supplanted the admittedly superior theory of the great Frenchman. Gauss in addition to his theory furnished an essentially useful and elegant method for performing the neces-

sary numerical calculations, while Laplace left this decidedly important aspect out of consideration altogether. It remained in reality to Charlier to furnish the Laplacean doctrine with a practical method for computing the various statistical parameters. And in the meantime the Gaussian methods reigned supreme while Laplace's great work was neglected.

The careful reader will readily notice that in the treatment of frequency curves I have allowed the *semi-invariants*, originally introduced in the theory of statistics by Thiele, to occupy a central position. In my opinion the semi-invariants represent a more powerful tool than the method of moments. I have also tried to rescue from oblivion the important and original memoir by the Danish actuary, Gram, and give to him and the French mathematician, Hermite, their due recognition as the earliest investigators of skew frequency distributions. Gram was perhaps the first investigator to make proper use of the orthogonal functional properties of the Laplacean normal frequency curve and its derivatives. By means of an application of the orthogonal properties of the Hermite polynomials and their close relation to the theory of integral equations, the whole theory of frequency distribution can be presented in a decidedly compact form; and I deem no apology necessary for having introduced in my treatment of frequency curves some of the more elementary theorems of integral equations, that youngest branch of higher analysis, which at present occupies a central position in advanced mathematics.

The most recent investigations along those lines have been made by the Swedish astronomer, Charlier, and his disciples, Jõrgensen and Wicksell. Unfortunately these investigations have hitherto not received adequate and systematic treatment in English and American texts on statistics, and it is my hope that the following pages may be of service in opening the eyes of English speaking statisticians to the practical utility of these methods.

The examples have all been selected so as to give a complete and detailed illustration of the application of the theory to essentially practical problems. I have, on the other hand, purposely refrained from giving the customary *exercises*, so-called, usually found in statistical texts, especially those in German and English.

Although I have been a close student of and have read most of the published statistical text-books in about seven languages for the last ten years, I regret to state that I have found little or no

practical value in such trick exercises, which as a rule have but slight relation to problems occurring in daily life.

Since the appearance of the first edition of this book in 1916 a number of excellent statistical texts have been issued. Among these I may mention a new edition of Yule's well-known elementary text, a greatly enlarged edition of Bowley's *Elements of Statistics*, the new treatise by Caradog Jones, an enlarged German translation of Charlier's *Grunddragen*, a very lucid Swedish text by Wicksell, the scholarly and broadly planned *Statistikens Teori i Grundrids* (in Danish) by Westergaard, and last but not least, the thesis by Jörgensen, *Frekvensflader og Korrelation.*[1]

Although an extended residence in the United States has perhaps improved my barbaric Dano-English, I fear that I must still apologize to the reader for my shortcomings in rhetoric and grammar. Most of the serious defects have, I hope, been overcome by the diligent efforts of my co-editor and translator, Miss C. Dickson, mathematical assistant in the department of Development and Research of the American Telephone and Telegraph Company. Miss Dickson's work has indeed been much beyond that of mere translation. Her knowledge of the mathematical theory of probabilities has enabled her to suggest to me several improvements in my Danish notes.

I am also under great obligations to a number of friends and colleagues who have assisted me in the preparation of this volume. I am especially indebted to Mr. E. C. Molina, the well-known probability expert of the American Telephone and Telegraph Company. Mr. Molina's extensive knowledge of the works of the old French masters, especially of those of Laplace, has been of the greatest value to me, and I can truthfully say that I have nowhere met a mathematician so thoroughly acquainted with the intricacies of the *Théorie Analytique* as Mr. Molina.

My thanks are also due to Mr. F. L. Hoffman, the Statistician of the Prudential Insurance Company, for the interest he took in my work along those lines while I was employed as a computer in his department. To Messrs. M. C. Rorty and D. R. Belcher of the American Telephone and Telegraph Company, I beg leave to

[1] As a pure probability text we may mention G. Castelnuovo's, *Calcolo delle Probabilita* (Milano, 1919), as an exceptionally lucid and rigorous treatise. The recently issued *Treatise on Probability* by J. M. Keynes is briefly discussed in paragraph 138 of this book. *A. F.*

express my best thanks for their kind advice and encouragement in the preparation of this volume.

It is indeed impossible to adequately express in a mere formal preface my obligations to Mr. Rorty in this matter. His introductory note I regard as one of the highest rewards I have received in this field of endeavor where one must usually be content with the appreciation of one's peers. In this connection it is of interest to note that Mr. Rorty is the pioneer investigator in the application of the mathematical theory of probabilities to telephone engineering, which has been further developed in recent years by Molina of America, Erlang and Johannsen of Denmark, Holm of Sweden, Odell and Grinsted of Great Britain. The pioneer work by Mr. Rorty in this eminently practical field antedates the earliest work by Erlang in *Tidsskrift for Matematik* by nearly five years.

Last, but not least, I wish to convey my sincerest thanks to my Scandinavian compatriots, Westergaard, Charlier, Jörgensen, Wicksell and Guldberg from whose works I have drawn so freely. To these gentlemen and to the works of the late Messrs. Gram and Thiele of Copenhagen I really owe anything of value which may be contained in this work.

ARNE FISHER.

New York,
April, 1922.

INTRODUCTORY NOTE TO THE FIRST EDITION.

I feel it a great honor to have been asked by my friend and colleague, Mr. Arne Fisher, of the Equitable Life Assurance Society of the United States, to write an introductory note to what appears to me the finest book as yet compiled in the English language on the subject of which it treats. As an Examiner myself in Statistical Method for a British Colonial Government, it has been to me a heart-breaking experience, when implored by intending candidates for examination to recommend a text-book dealing with Mr. Fisher's subject matter, that it has heretofore been impossible for me to recommend one in the English language which covers the whole of the ground. Until comparatively recent years the case was even worse. While in French, in Italian, in German, in Danish, and in Dutch, scientific works on statistics were available galore, the dearth of such literature in the English language was little short of a national or racial scandal. With such works as those of Yule and Bowley, in recent years, there has been some possibility for the English-speaking student to acquire part of the knowledge needed. But it is hardly necessary to point out what a very large amount of new ground is covered by Mr. Fisher's new book as compared with such works as I have referred to.

Despite my professional connection with statistical and actuarial work of a technical character my own personal interest in Mr. Fisher's book is concentrated principally on the metaphysical basis of the Probability-theory, and it is with regard to this aspect of the subject alone that I feel qualified to comment on his achievement. With all the controversy that has gone on through many decades among metaphysicians and among writers on logic interested especially in the bases of the theories of probability and induction, between the pure empiricists of the type of J. S. Mill and John Venn (at all events in the earliest edition of his work) on the one hand, and the (partly) *à priori* theorists who base their doctrine on the foundation of Laplace on the other hand, it has

been a source of intense satisfaction to me, as in the main a disciple of the latter group of theorists, to note the masterly way in which Mr. Arne Fisher disentangles the issues which arise in the keen and sometimes almost embittered controversy between these two schools of thought. It has always seemed to the present writer as if the very foundations of Epistemology were involved in this controversy. The impossibility of deriving the corpus of human knowledge exclusively from empirical data by any logically valid process—an impossibility which led Immanuel Kant to the creation of his epoch-making philosophical system—is hardly anywhere made more evident than in what seems to the present writer the unsuccessful effort of thinkers like John Venn to derive from such purely empirical data the entire Theory of Probability. The logical fallacy of the process is analogous to that perpetrated by John Stuart Mill in endeavoring to base the Law of Causality on what he termed an "*inductio per simplicem enumerationem.*" Probably there is nowhere a more trenchant and conclusive exposure of the unsoundness of this point of view, than in the Right Honorable Arthur James Balfour's monumental work "A Defense of Philosophic Doubt." It is therefore satisfactory to find that Mr. Fisher emphasizes, quite at the beginning of his treatise, that an *à priori* foundation for "Probability" judgments is indispensable.

Hardly less gratifying, from the metaphysical point of view, is Mr. Fisher's treatment of the celebrated *quaestio vexata* of Inverse Probabilities and his qualified vindication of Bayes' Rule against its modern detractors.

Aside altogether from metaphysics, it is particularly satisfactory to note the full and clear way in which the author treats the Lexian Theory of Dispersion and of the "Stability" of statistical series and the extension of this theory by recent Scandinavian and Russian investigators,—a branch of the science which has till the appearance of this new work not been adequately covered in English text-books.

It may of course be a moot question whether the preference given by our author to Charlier's method of treating "Frequency Curves" over the method of Professor Karl Pearson is well advised. But whatever the experts' verdict may be on debatable

questions like these, the scientific world is to be congratulated on Mr. Fisher's presentment of a new and sound point of view, and he emphatically is to be congratulated on the production of a text-book which for many years to come will be invaluable both to students and to his confrères who are engaged in extending the boundaries of this fascinating science.

F. W. FRANKLAND,
Member of the Actuarial Society of America,
Fellow of the Institute of Actuaries of Great
Britain and Ireland, and Fellow of the
Royal Statistical Society of London.

NEW YORK,
October 1, 1915.

PREFACE TO THE FIRST EDITION.

"Probability" has long ago ceased to be a mere theory of games of chance and is everywhere, especially on the continent, regarded as one of the most important branches of applied mathematics. This is proven by the increasing number of standard text-books in French, German, Italian, Scandinavian and Russian which have appeared during the last ten years. During this time the research work in the theory of probabilities has received a new impetus through the labors of the English biometricians under the leadership of Pearson, the Scandinavian statisticians Westergaard, Charlier and Kiær, the German statistical school under Lexis, and the brilliant investigations of the Russian school of statisticians.

Each group of these investigations seems, however, to have moved along its own particular lines. The English schools have mostly limited their investigations to the field of biology as published in the extensive memoirs in the highly specialized journal, *Biometrika.* The Scandinavian scholars have produced researches of a more general character, but most of these researches are unfortunately contained in Scandinavian scientific journals and are for this reason out of reach to the great majority of readers who are not familiar with any of the allied Scandinavian languages. This applies in a still greater degree to the Russians. German scholars of the Lexis school have also contributed important memoirs, but strangely enough their researches are little known in this country or in England, a fact which is emphasized through the belated English discussion on the theory of dispersion as developed by Lexis and his disciples. The same can also be said with regard to the Italian statisticians.

In the present work I have attempted to treat all these modern researches from a common point of view, based upon the mathematical principles as contained in the immortal work of the great Laplace, "Théorie analytique des Probabilités," a work which despite its age remains the most important contribution to the

theory of probabilities to our present day. Charlier has rightly observed that the modern statistical methods may be based upon a few condensed rules contained in the great work of Laplace. This holds true despite the fact that many modern English writers of late have shown a certain distrust, not to say actual hostility, towards the so-called mathematical probabilities as defined by the French savant, and have in their place adopted the purely empirical probability ratios as defined by Mill, Venn and Chrystal. It is quite true that it is possible to build a consistent theory of such ratios, as for an instance is done by the Danish astronomer and actuary, Thiele. The theory, however, then becomes purely a theory of observations in which the theory of probability takes a secondary place. The distrust in the so-called mathematical a priori probabilities of Laplace I believe, however, to be unfounded, and the criticism to which that particular kind of probabilities is subjected by a few of the modern English writers is, I believe, due to a misapprehension of the true nature of the Bernoullian Theorem. This renowned theorem remains to-day the cornerstone of the theory of statistics, and upon it I have based the most important chapters of the present work. Following the beautiful investigations of Tschebycheff and Pizetti in their proofs of Bernoulli's Theorem and the closely related theorem of large numbers by Poisson I have adopted the methods of the Swedish astronomer and statistician, Charlier, in the discussion of the Lexian dispersion theory.

The theory of frequency curves is treated from various points of view. I have first given a short historical introduction to the various investigations of the law of errors. The Gaussian normal curve of error was by the older school of statisticians held to be sufficient to represent all statistical frequencies, and actual observed deviations from the normal curve were attributed to the limited number of observations. Through the original memoirs of Lexis and the investigations of Thiele the fallacy of such a dogmatic belief was finally shown. The researches of Thiele, and later of Pearson, developed later the theory of skew curves of error. As recently as 1905 Charlier finally showed that the whole theory of errors or frequency curves may be brought back to the principles of Laplace. I have treated this

subject by the methods of both Pearson and Charlier, although I have given the methods of the latter a predominant place, because of their easy and simple application in the practical computations required by statistical work. The mathematical theory of correlation, which is treated in an elementary manner only, is based upon the same principles.

The statistical examples serve as illustrations of the theory, and it will be noted that it is possible to solve all the important statistical problems presenting themselves in daily work on the basis of a theory of mathematical probabilities instead of on a direct theory of statistical methods. I have here again followed Charlier in dividing all statistical problems into two distinct groups, namely, the homograde and the heterograde groups.

In treating the philosophical side of the subject I have naturally not gone into much detail. However, I have tried to emphasize the two diametrically opposite standpoints, namely the principle of what von Kries has called the principle of "cogent reason," and the principle which Boole has aptly termed "the equal distribution of ignorance." These two principles are clearly illustrated in the case of the so-called inverse probabilities. As far as pure theory is concerned, the theory of "inverse probabilities" is rigorous enough. It is only when making practical applications of the rule of inverse probabilities (the so-called Bayes' Rule) that many writers have made a fatal mistake by tacitly assuming the principle of "insufficient reason" as the only true rule of computation. This leads to paradoxical results as illustrated by the practical problem from the region of actuarial science in Chapter VI in this book.

In a work of this character I have naturally made an extended use of the higher mathematical analysis. However, the reader who is not versed in these higher methods need not feel alarmed on this account, as the elementary chapters are arranged in such a way that the more difficult paragraphs may be left out. I have in fact divided the treatise into two separate parts. The first part embraces the mathematical probabilities proper and their applications to homograde statistical series. This part, I think, constitutes what is usually given as a course in vital statistics in many American colleges. I hardly deem it worth while to give a

detailed discussion on the collection and arrangement of the statistical data as to various frequency distributions. The mere graphical and serial representation of frequency functions by means of histographs and frequency columns is so simple and evident that a detailed description seems superfluous. The fitting of the various curves to analytical formulas and the determination of the various parameters seem to me of much greater importance. The theory of curve fitting which is treated in the second volume is founded upon a more advanced mathematical analysis and is for this reason out of reach to the average American student who desires to learn only the rudiments of modern statistical methods. Practical statisticians, on the other hand, will derive much benefit from these higher methods. It is a fact generally noted in mathematics that the practical application of a difficult theory is much simpler than that of a more elementary theory. This is amply proven by the appearance of an excellent little Scandinavian brochure by Charlier: "Grunddragen af den matematiske Statistikken." ("Rudiments of Mathematical Statistics.") I have always attempted to adapt theory to actual practical problems and requirements rather than to give a purely mathematical abstract discussion. In fact it has been my aim to present a theory of probabilities as developed in recent years which would prove of value to the practical statistician, the actuary, the biologist, the engineer and the medical man, as well as to the student who studies mathematics for the sake of mathematics alone.

The nucleus of this work consisted of a number of notes written in Danish on various aspects of the theory of probabilities, collected from a great number of mathematical, philosophical and economic writings in various languages. At the suggestion of my former esteemed chief, Mr. H. W. Robertson, F.A.S., Assistant Actuary of the Equitable Life Assurance Society of the United States, I was encouraged to collect these fragmentary notes in systematic form. The rendering in English was done by myself personally with the assistance of Mr. W. Bonynge. With his assistance most of the idiomatic errors due to my barbaric Dano-English have been eliminated. The notes stand, however, in the main as a faithful reproduction of my original

English copy. Although the resulting "Dano-English" may have its great shortcomings as to rhetoric and grammar, I hope to have succeeded in expressing what I wanted to say in such a manner that my possible readers may follow me without difficulty.

I gladly take the opportunity of expressing my thanks to a number of friends and colleagues who in various ways have assisted me in the preparation of this work. My most grateful thanks are due to Mr. F. W. Frankland, Mr. H. W. Robertson and Mr. Wm. Bonynge not only for reading the manuscript and most of the proofs, but also for the friendly help and encouragement in the completion of this volume. The introductory note by Mr. Frankland, coming from the pen of a scholar who for the most of a life-time has worked with statistical-mathematical subjects and who has taken a special interest in the philosophical and metaphysical aspects of the probability theory, I regard as one of the strong points of the book. My debts to Messrs. Frankland and Robertson as well as to Dr. W. Strong, Associate Actuary of the Mutual Life Insurance Company, are indeed of such a nature that they cannot be expressed in a formal preface. My thanks are also due to Mr. A. Pettigrew in correcting the first rough draught of the first three chapters at a time when my knowledge of English was most rudimentary, to Mr. M. Dawson, Consulting Actuary, and Mr. R. Henderson, Actuary of the Equitable Life, for reading a few chapters in manuscript and making certain critical suggestions, to Professors C. Grove and W. Fite, of Columbia University, for numerous technical hints in the working out of various mathematical formulas in Chapter VI, to Miss G. Morse, librarian of the Equitable Library, in the search of certain bibliographical material. Last but not least I wish to express my sincerest thanks to several of my Scandinavian compatriots for allowing me to quote and use their researches on various statistical subjects. I want in this connection especially to mention Professor Charlier, of Lund, and Professors Westergaard and Johannsen, of Copenhagen.

To The Macmillan Company and The New Era Printing Company I beg leave to convey my sincere appreciation of their very courteous and accommodating attitude in the manufacture of

this work. Their spirit has been far from commercial in this—from a pure business standpoint—somewhat doubtful undertaking.

ARNE FISHER.

NEW YORK,
October, 1915.

TABLE OF CONTENTS.

PART I.

MATHEMATICAL PROBABILITIES AND HOMOGRADE STATISTICS.

CHAPTER I.

Introduction: General Principles and Philosophical Aspects.

CHAPTER II.

Historical and Bibliographical Notes.

CHAPTER III.

The Mathematical Theory of Probabilities.

CHAPTER IV.

The Addition and Multiplication Theorems in Probabilities.

CHAPTER IX.

Law of Large Numbers. Mathematical Deduction.

CHAPTER X.

The Theory of Dispersion and the Criterions of Lexis and Charlier.

CHAPTER XI.

Application to Games of Chance and Statistical Problems.

CHAPTER XII.

Continuation of the Application of the Theory of Probabilities to Homograde Statistical Series.

PART II.

FREQUENCY CURVES AND HETEROGRADE STATISTICS

CHAPTER XIII.

The Theory of Errors and Frequency Curves and Its Application to Statistical Series. General Remarks.

CHAPTER XIV.

The Mathematical Theory of Frequency Curves.

PART III.

PRACTICAL APPLICATIONS OF THE THEORY.

CHAPTER XV.

The Numerical Determination of the Parameters.

CHAPTER XVI.

Logarithmically Transformed Frequency Functions.

CHAPTER XVII.

Frequency Curves and their Relation to the Bernoullian Series.

CHAPTER XVIII.

Poisson-Charlier Frequency Curves for Integral Variates.

PART I

MATHEMATICAL PROBABILITIES AND HOMOGRADE STATISTICS

CHAPTER I.

INTRODUCTION: GENERAL PRINCIPLES AND PHILOSOPHICAL ASPECTS.

1. Methods of Attack.—The subject of the theory of probabilities may be attacked in two different ways, namely in a philosophical, and in a mathematical manner. At first the subject originated as isolated mathematical problems from games of chance. The pioneer writers on probability such as Cardano, Galileo, Pascal, Fermat, and Huyghens treated it in this way. The famous Bernoulli was, perhaps, the first to view the subject from the philosopher's point of view. Laplace wrote his well-known "Essai Philosophique des Probabilités," wherein he terms the whole science of probability as the application of common sense. During the last thirty years numerous eminent philosophical scholars such as Mill, Venn, and Keynes of England, Bertrand and Poincaré of France, Sigwart, von Kries and Lange of Germany, Kroman of Denmark, and several Russian scholars have written on the philosophical aspect.

In the ordinary presentation of the elements of the theory of probability as found in most English text-books, the treatment is wholly mathematical. The student is given the definition of a mathematical probability and the elementary theorems are then proved. We shall, in the following chapter, depart from this rule and first view the subject, briefly, from a philosophical standpoint. What the student may thus lose in time we hope he may gain in obtaining a broader view of the fundamental principles underlying our science. At the same time, the reader who is unacquainted with the science of philosophy or pure logic, need not feel alarmed, since not even the most elementary knowledge of the principles of formal logic is required for the understanding of the following chapter.

2. Law of Causality.—In a great treatise on the Chinese civilization, Oscar Peschel, the German geographer and philosopher, makes the following remarks: "Since our intellectual awakening, since we have appeared on the arena of history as the creators

and guardians of the treasures of culture, we have sought after only one thing, of the presence of which the Chinese had no idea, and for which they would give hardly a bowl of rice. This invisible thing we call causality. We have admired a vast number of Chinese inventions, but even if we seek through their huge treasures of philosophical writing we are not indebted to them for a single theory or a single glance into the relation between cause and effect."

The law of causality may be stated broadly as follows: Everything that happens, and everything that exists, necessarily happens or exists as the consequence of a previous state of things. This law cannot be proven. It must be taken, a priori, as an axiom; but once accepted as a truth it does away with the belief of a capricious ruling power, and even if the strongest disbeliever of the law may deny its truth in theory he invariably applies it in practice during his daily occupation in life.

All future human activity is more or less influenced by past and present conditions. Modern historical writings, as for instance the works of the brilliant Italian historian, Ferrero, always seek to connect past events with present social and economic conditions. Likewise great and constructive statesmen in trying to shape the destinies of nations always reckon with past and present events and conditions. We often hear the term, "a man with foresight," applied to leading financiers and statesmen. This does not mean that such men are gifted with a vision of the future, but simply that they, with a detailed and thorough knowledge of past and present events, associated with the particular undertaking in which they are interested, have drawn conclusions in regard to a future state of affairs. For example, when the Canadian Pacific officials, in the early eighties, chose Vancouver as the western terminal for the transcontinental railroad, at a time when practically the whole site of the present metropolis of western Canada was only a vast timber tract, they realized that the conditions then prevailing on this particular spot—the excellent shipping facilities, the favorable location in regard to the Oriental trade, and the natural wealth of the surrounding country—would bring forth a great city, and their predictions came true.

Predictions with regard to the future must be taken seriously only when they are based upon a thorough knowledge of past and present events and conditions. Prophecies, taken in a purely biblical sense of the term and viewed from the law of causality, are mere guesses which may come true and may not. A prophet can hardly be called more than a successful guesser. Whether there have been persons gifted with a purely prophetic vision is a question which must be left to the theologians to wrangle over.

3. **Hypothetical Judgments.**—Any person with ordinary intellectual faculties may, however, predict certain future events with absolute certainty by a simple application of the principle of hypothetical judgment. The typical form of the hypothetical judgment is as follows: If a certain condition exists, or if a certain event takes place then another definite event will surely follow. Or if A exists B will invariably follow.

Mathematical theorems are examples of hypothetical judgments. Thus in the geometry of the plane we start with certain ideas (axioms) about the line and plane. From these axioms we then deduce the theorems by mere hypothetical judgments. Thus in the Euclidian geometry we find the axiom of parallel lines, which assumes that through a point only one line can be drawn parallel to another given line, and from this assumption we then deduce the theorem that the sum of the angles in a triangle is 180°. But it must be borne in mind that this proof is valid only on the assumption of the actual existence of such lines. If we could prove directly by logical reasoning or by actual measurement, that the sum of the angles in any triangle is equal to 180°, then we would be able to prove the above theorem, the so-called "hole in geometry," independently of the axiom of parallel lines.

A Russian mathematician, Lobatschewsky, on the other hand, assumed that through a single point an infinite number of parallels might be drawn to a previously given line, and from this assumption he built up a complete and valid geometry of his own. Still another mathematician, Riemann, assumed that no lines were parallel to each other, and from this produced a perfectly valid surface geometry of the sphere.

As examples of hypothetical judgment we have the two following well-known theorems from elementary geometry and algebra. If one of the angles of a triangle is divided into two parts, then the line of division intersects the opposite side. If a decadian number is divided by 5 there is no remainder from the division.

In natural science, hypothetical judgments are founded on certain occurrences (phenomena) which, without exception, have taken place in the same manner, as shown by repeated observations. The statement that a suspended body will fall when its support is removed is a hypothetical judgment derived from actual experience and observation.

4. Hypothetical Disjunctive Judgments.—In hypothetical judgments we are always able to associate cause and effect. It happens frequently, however, that our knowledge of a certain complex of present conditions and actions is such that we are not able to tell beforehand the resulting consequences or effects of such conditions and actions, but are able to state only that either an event E_1 or an event E_2, etc., or an event E_n will happen. This represents a hypothetical disjunctive judgment whose typical form is: If A exists either E_1, E_2, E_3, $\cdots$ or E_n will happen.

If we take a die, *i. e.*, a homogeneous cube whose faces are marked with the numbers from one to six, and make an ordinary throw, we are not able to tell beforehand which side will turn up. True, we have here again a previous state of things, but the conditions do not allow such a simple analysis as the cases we have hitherto considered under the purely hypothetical judgment. Here a multitude of causes influence the final result—the weight and centre of gravity of the die, the infinite number of possible movements of the hand which throws the die, the force of contact with which the die strikes the table, the friction, etc. All these causes are so complex that our minds are not afforded an opportunity to grasp and distinguish the impulses that determine the fall of the die. In other words we are not able to say, a priori, which face will appear. We only know for certain that either 1, 2, 3, 4, 5, or 6 will appear. If a line is drawn through the vertex of a triangle, it either intersects the opposite side or it does not. If a number is divided by 5 the division either gives

only an integral number or leaves a remainder. If an opening is made in the wall of a vessel partly filled with water, then either the water escapes or remains in the vessel. All the above cases are examples of hypothetical disjunctive judgments.

The four cases show, however, a common characteristic. They all have a certain partial domain, where one of the mutually exclusive events is certain to happen, while the other partial domain will bring forth the other event, and the total area of action embraces both events. Taking the triangle, we notice that the lines may pass through all the points inside of an angle of 360°, but only the lines falling inside the internal vertical angle, φ, of the triangle will produce the event in question, namely the line intersecting the opposite side. There will be an outflow from the vessel only if the hole is made in that part of the wall which is touched by the fluid.

All problems do not allow of such simple analysis, however, as will be seen from the following example. Suppose we have an urn containing 1 white and 2 black balls and let a person draw one from the urn. The hypothetical disjunctive judgment immediately tells us that the ball will be either black or white, but the particular domain of each event cannot be limited to the fixed border lines of the former examples. Any one of the balls may occupy an infinite number of positions, and furthermore we may imagine an infinite number of movements of the hand which draws the ball, each movement being associated with a particular point of position of the ball in the urn. If we now assume each of the three balls to have occupied all possible positions in the urn, each point of position being associated with its proper movement of the hand, it is readily seen that a black ball will be encountered twice as often as a white ball in a particular point of position in the urn, and for this reason any particular movement of the hand which leads to this point of position grasps a black ball twice as often as a white ball.

5. General Definition of the Probability of an Event.—All the above examples have shown the following characteristics:

(1) A total general region or area of action in which all actions may take place, this total area being associated with all possible events.

(2) A limited special domain in which the associated actions produce a special event only.

If these areas and domains, as in the above cases, are of such a nature that they allow a purely quantitative determination, they may be treated by mathematical analysis. We define now, without entering further into its particular logical significance, the ratio of the second special and limited domain to the first total region or area as the probability of the happening of the event, E, associated with domain No. 2.

We must, however, hasten to remark that it is only in a comparatively few cases that we are able, a priori, to make such a segregation of domains of actions. This may be possible in purely abstract examples, as for instance in the example of the division of the decadian number by 5. But in all cases where organic life enters as a dominant factor we are unable to make such sharp distinctions. If we were asked to determine the probability of an x-year-old person being alive one year from now, we should be able to form the hypothetical disjunctive judgment: An x-year-old person will be either alive or dead one year from now. But a further segregation into special domains as was the case with the balls in the urn is not possible. Many extremely complex causes enter into such a determination; the health of the particular person, the surroundings, the daily life, the climate, the social conditions, etc. Our only recourse in such cases is to actual observation. By observing a large number of persons of the same age, x, we may, in a purely empirical way, determine the rate of death or survival. Such a determination of an unknown probability is called an empirical probability. An empirical probability is thus a probability, into the determination of which actual experience has entered as a dominant factor.

6. Equally Likely Cases.— The main difficulty, in the application of the above definition of probability, lies in the determination of the question whether all the events or cases taking place in the general area of action may be regarded as equally likely or not. Two diametrically opposite views have here been brought forward by writers on probabilities. One view is based upon the principle which in logic is known as the principle of

"insufficient reason," while the other view is based upon the principle of "cogent reason." The classical writers on the theory of probability, such as Jacob Bernoulli and Laplace, base the theory on the principle of insufficient reason exclusively. Thus Bernoulli declares the six possible cases by the throw of a die to be equally likely, since "on account of the equal form of all the faces and on account of the homogeneous structure and equally arranged weight of the die, there is no reason to assume that any face should turn up in preference to any other." In one place Laplace says that the possible cases are "cases of which we are equally ignorant," and in another place, "we have no reason to believe any particular case should happen in preference to any other."

The opposite view, based on the principle of cogent reason, has been strongly endorsed in an admirable little treatise by the German scholar, Johannes von Kries.[1] Von Kries requires, first of all, as the main essential in a logical theory of probability, that "the arrangement of the equally likely cases must have a cogent reason and not be subject to arbitrary conditions."

In several illustrative examples, von Kries shows how the principle of insufficient reason may lead to different and paradoxical results. The following example will illustrate the main points in von Kries's criticism. Suppose we be given the following problem: Determine the probability of the existence of human beings on the planet Mars. By applying the first mentioned principle our reasoning would be as follows: We have no more reason to assume the actual existence of man on the planet than the complete absence. Hence the probability for the non-existence of a human being, is equal to $\frac{1}{2}$. Next we ask for the probability of the presence or non-presence of another earthly mammal, say the elephant. The answer is the same, $\frac{1}{2}$. Now the probability for the absence of both man and elephant on the planet is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.[2] The probability for the absence of a third mammal, the horse, is also $\frac{1}{2}$, or the probability for the absence of man, elephant, and horse is equal to $(\frac{1}{2})^3 = \frac{1}{8}$. Proceeding in the same manner for all mammals we obtain a very small proba-

[1] "Die Principien der Wahrscheinlichkeitsrechnung." Berlin, 1886.

[2] See the chapter on multiplication of probabilities.

bility for the complete absence of all mammals on Mars, or a very large probability, almost equal to certainty, that the planet harbors at least one mammal known on our planet, an answer which certainly does not seem plausible. But we might as well have put the question from the start: what is the probability of the existence or absence of any one earthly mammal on Mars? The principle of insufficient reason when applied directly would here give the answer $\frac{1}{2}$, while when applied in an indirect manner the same method gave an answer very near to certainty.

An urn is known to contain white and black balls, but the number of the balls of the two different colors is unknown. What is the probability of drawing a white ball? The principle of insufficient reason gives us readily the answer: $\frac{1}{2}$, while the principle of cogent reason would give the same answer only if it were known a priori that there were equal numbers of balls of each color in the urn before the drawing took place. Since this knowledge is not present a priori, we are not able to give any answer, and the problem is considered outside the domain of probabilities. There is no doubt that the principle advocated by von Kries is the only logical one to apply, and a recent treatise on the theory of probability by Professor Bruhns of Leipzig[1] also gives the principle of cogent reason the most prominent place. On the other hand it must be admitted that if the principle was to be followed consistently in its very extreme it would of course exclude many problems now found in treatises on probability and limit the application of our theory considerably in scope. Still, however, we must agree with von Kries that it seems very foolhardy to assign cases of which we are absolutely in the dark, as being equally likely to occur. This very principle of insufficient reason is in very high degree responsible for the somewhat absurd answers to questions on the so-called "inverse probabilities," a name which in itself is a great misnomer. We shall later in the chapter on "a posteriori" probabilities discuss this question in detail. At present we shall only warn the student not to judge cases of which he has no knowledge whatsoever to be equally likely to occur. The old rule "experience is the best teacher" holds here, as everywhere else.

[1] "Kollektivmasslehre and Wahrscheinlichkeitsrechnung," Leipzig, 1903.

7. Objective and Subjective Probabilities.—In this connection it is interesting to note the lucid remarks by the Danish statistician, Westergaard. "By every well arranged game of chance, by lotteries, dice, etc.," Westergaard says, "everything is arranged in such a way that the causes influencing each draw or throw remain constant as far as possible. The balls are of the same size, of the same wood, and have the same density; they are carefully mixed and each ball is thus apparently subject to the influences of the same causes. However, this is not so. Despite all our efforts the balls are different. It is impossible that they are of exactly mathematically spherical form. Each ball has its special deviation from the mathematical sphere, its special size and weight. No ball is absolutely similar to any one of the others. It is also impossible that they may be situated in the same manner in the bag. In short there is a multitude of apparently insignificant differences which determine that a certain definite ball and none of the other balls may be drawn from the bag. If such inequalities did not exist one of two things would happen. Either all balls would turn up simultaneously or also they would all remain in the bag. Many of these numerous causes are so small that they perhaps are invisible to the naked eye and completely escape all calculations, but by mutual action they may nevertheless produce a visible result."

It thus appears that a rigorous application of the principle of cogent reason seems impossible. However, a compromise between this principle and that of the principle of insufficient reason may be effected by the following definition of equally possible cases, viz.: *Equally possible cases are such cases in which we, after an exhaustive analysis of the physical laws underlying the structure of the complex of causes influencing the special event, are led to assume that no particular case will occur in preference to any other.* True, this definition introduces a certain subjective element and may therefore be criticized by those readers who wish to make the whole theory of probabilities purely objective. Yet it seems to me preferable to the strict application of the principle of equal distribution of ignorance. Take again the question of the probability of the existence of human beings on the planet Mars. The principle of equal distribution of ignorance

readily gives us without further ado the answer $\frac{1}{2}$. Modern astrophysical researches have, however, verified physical conditions on the planet which make the presence of organic life quite possible, and according to such an eminent authority as Mr. Lowell, perhaps absolutely certain. Yet these physical investigations are as yet not sufficiently complete, and not in such a form that they may be subjected to a purely quantitative analysis as far as the theory of probabilities is concerned. Viewed from the standpoint of the principle of cogent reason any attempt to determine the numerical value of the above probability must therefore be put aside as futile. This result, negative as it is, seems, however, preferable to the absolute guess of $\frac{1}{2}$ as the probability.

CHAPTER II.

HISTORICAL AND BIBLIOGRAPHICAL NOTES.

8. Pioneer Writers.—The first attempt to define the measure of a probability of a future event is credited to the Greek philosopher, Aristotle. Aristotle calls an event probable when the majority, or at least the majority of the most intellectual persons, deem it likely to happen. This definition, although not allowing a purely quantitative measurement, makes use of a subjective judgment.

The first really mathematical treatment of chance, however, is given by the two Italian mathematicians, Cardano and Galileo, who both solved several problems relating to the game of dice. Cardano, aside from his mathematical occupation, was also a professional gambler and had evidently noticed that in all kinds of gambling houses cheating was often resorted to. In order that the gamester might be fortified against such cheating practices, Cardano wrote a little treatise on gambling wherein he discussed several mathematical questions connected with the different games of dice as played in the Italian gambling houses at that time. Galileo, although not a professional gambler, was often consulted by a certain Italian nobleman on several problems relating to the game of dice, and fortunately the great scholar has left some of his investigations in a short memoir. In the same manner the two great French mathematicians, Pascal and Fermat, were often asked by a professional gamester, the chevalier de Mere, to apply their mathematical skill to the solution of different gambling problems. It was this kind of investigation which probably led Pascal to the discovery of the arithmetical triangle, and the first rudiments of the combinatorial analysis, which had its origin in probability problems, and which later evolved into an independent branch of mathematical analysis.

One of the earliest works from the illustrious Dutch physicist, Huyghens, is a small pamphlet entitled "de Ratiociniis in Ludo Aleæ," printed in Leyden in the year 1657. Huyghens' tract is

the first attempt of a systematic treatment of the subject. The famous Leibnitz also wrote on chance. His first reference to a mathematical probability is perhaps in a letter to the philosopher, Wolff, wherein he discusses the summation of the infinite series $1 - 1 + 1 - 1 + \cdots$. Besides he solved several problems.

9. Bernoulli, de Moivre and Bayes.—The first extensive treatise on the theory as a whole is from the hand of the famous Jacob Bernoulli. Bernoulli's book, "Ars Conjectandi," marks a revolution in the whole theory of chance. The author treats the subject from the mathematical as well as from a philosophical point of view, and shows the manifold applications of the new science to practical problems. Among other important theorems we here find the famous proposition which has become known as the Bernoulli Theorem in the mathematical theory of probabilities. Bernoulli's work has recently been translated from the Latin into German,[1] and a student who is interested in the whole theory of probability should not fail to read this masterly work.

The English mathematicians were the next to carry on the investigations. Abraham de Moivre, a French Huguenot, and one of the most remarkable mathematicians of his time, wrote the first English treatise on probabilities.[2] This book was certainly a worthy product of the masterful mind of its author, and may, even today, be read with useful results, although the method of demonstration often appears lengthy to the student who is accustomed to the powerful tools of modern analysis. The high esteem in which the work by de Moivre is held by modern writers, is proven by the fact that E. Czuber, the eminent Austrian mathematician and actuary, so recently as two years ago translated the book into German. A certain problem (see Chap. IV) still goes under the name of "The Problem of de Moivre" in the modern literature on probability. A contemporary of de Moivre, Stirling, contributed also to the new branch of mathematics, and his name also is immortalized in the theory of probability by the formula which bears his name, and by which we are able to express large factorials to a very accurate degree of approximation. The third important English contributor is

[1] Ars Conjectandi, Ostwald's Klassiker No. 108, Leipzig, 1901.

[2] de Moivre: "The Doctrine of Chances," London, 1781.

the Oxford clergyman, T. Bayes. Bayes' treatise, which was published after his death by Price, in *Philosophical Transactions* for 1764, deals with the determination of the a posteriori probabilities, and marks a very important stepping stone in our whole theory. Unfortunately the rule known as Bayes' Rule has been applied very carelessly, and that mostly by some of Bayes' own countrymen; so the whole theory of Bayes has been repudiated by certain modern writers. A recent contribution by the Danish philosophical writer, Dr. Kroman, seems, however, to have cleared up all doubts on the subject, and to have given Bayes his proper credit.

10. Application to Statistical Data.—In the eighteenth century some of the most celebrated mathematicians investigated problems in the theory of probability. The birth of life assurance gave the whole theory an important application to social problems and the increasing desire for the collection of all kinds of statistical data by governmental bodies all over Europe gave the mathematicians some highly interesting material to which to apply their theories. No wonder, therefore, that we in this period find the names of some of the most illustrious mathematicians of that time, such as Daniel Bernoulli, Euler, Nicolas and John Bernoulli, Simpson, D'Alembert and Buffon, closely connected with the solution of problems in the theory of mathematical probabilities. We shall not attempt to give an account of the different works of these scientists, but shall only dwell briefly on the labors of Bernoulli and D'Alembert. In a memoir in the St. Petersburg Academy, Daniel Bernoulli is the first to discuss the so called St. Petersburg Problem, one of the most hotly debated in the whole realm of our science. We may here mention that this problem is today one of the main pillars in the economic treatment of value Bernoulli introduced in the discussion of the above mentioned problem the idea of the "moral expectation," which under slightly different names appears in nearly all standard writings on economics.

D'Alembert is especially remembered for the critical attitude he took towards the whole theory. Although one of the most brilliant thinkers of his age, the versatile Frenchman made some great blunders in his attempt to criticize the theories of chance.

Buffon's name is remembered because of the needle problem, and he may properly be called the father of the so-called "geometrical" or "local" probabilities.

11. Laplace and Modern Writers.—We now come to that resplendent genius in the investigation of the mathematical theory of chance, the immortal Laplace, who in his great work, "Théorie Analytique des Probabilités," gave the final mathematical treatment of the subject. This massive volume leaves nothing to be desired and is still today—more than one hundred years after its first publication—a most valuable mine of information and compares favorably with much more modern treatises. But like all mines, it requires to be mined and is by no means easy reading for a beginner. An elementary extract, "Essai Philosophique des Probabilités," containing the more elementary parts of Laplace's greater work and stripped of all mathematical formulas has recently appeared in an English translation.

Among later French works, Cournot's "Exposition de la Théorie des Chances et des Probabilités" (1843), treated the principal questions in the application of the theory to practical problems in sociology. In 1837 Poisson published his "Recherches sur les Probabilités" in which he for the first time proved the famous theorem which bears his name. Poisson and his Belgian contemporary, Quetelet, made extensive use of the theory in the treatment of statistical data.

Among the most recent French works, we mention especially Bertrand's "Calcul des Probabilités" (Paris, 1888), Poincaré's "Calcul des Probabilités" (Paris, 1896), and Borel's "Calcul des Probabilités" (Paris, 1901). We especially recommend Poincaré's brilliant little treatise to every student who masters the French language, as this book makes no departure from the lively and elucidating manner in which this able mathematical writer treated the numerous subjects on which he wrote during his long and brilliant career as a mathematician.

Of Russian writers, the mathematician, Tchebycheff, has given some extensive general theorems relating to the law of large numbers. Unfortunately Tchebycheff's writings are for the most part scattered in French, German, Scandinavian and

Russian journals, and thus are not easily accessible to the ordinary reader. A Russian artillery officer, Sabudski, has recently published a treatise on ballistics in German, wherein he extends the views formulated by Tchebycheff.

Of Scandinavian writers we mention T. N. Thiele, who probably was the first to publish a systematic treatise on skew curves.[1] An abridged edition of this very original work has recently been translated into English.[2] The Dane, Westergaard, is the author of the most extensive and thorough treatise on vital statistics which we possess at the present time. Westergaard's work has recently been translated into German,[3] and is strongly recommended to the student of vital statistics on account of his clear and attractive style of presenting this important subject.

The Swedish mathematicians Charlier and Gylden have published a series of memoirs in different Scandinavian journals and scientific transactions. We may also, in this category, mention the numerous small articles by the eminent Danish actuary, Dr. Gram.

While the German mathematicians in general are the most fertile writers on almost every branch of pure and applied mathematics, they have not shown much activity in the theory of mathematical probability except in the past ten years. But during that time there has appeared at least a dozen standard works in German. Among these, the lucid and terse treatise by E. Czuber, the Austrian actuary and mathematician, is especially attractive to the beginner on account of the systematic treatment of the whole subject.[4] A very original treatment is offered by H. Bruhns in his "Kollektivmasslehre und Wahrscheinlichkeitsrechnung" (Leipzig, 1903). Among the German works, we may also mention the book by Dr. Norman Herz in "Sammlung Schubert," and an excellent little work by Hack in the small pocket edition of "Sammlung Göschen." The theory of skew curves and correlation is presented by Lipps and Bruhns in extensive treatises.

[1] "Almindelig Iagttagelseslaere," Copenhagen, 1884.

[2] "Theory of Observations," London, 1903.

[3] "Mortalität und Morbilität," Jena, 1902.

[4] E. Czuber, "Wahrscheinlichkeitsrechnung," Leipzig, 1908 and 1910, 2 volumes.

We finally come to modern English writers on the subject. After the appearance of de Moivre's "Doctrine of Chances" the first work of importance was the book by de Morgan "An Essay on the Theory of Probabilities." The latest text-book is Whitworth's "Choice and Chance" (Oxford Press, 1904); but none of these works, although very excellent in their manner of treatment of the subject, comes up to the French, Scandinavian, and German text-books. Nevertheless, some of the most important contributions to the whole theory have been made by the English statisticians and mathematicians, Crofton, Pearson, and Edgeworth. Especially have frequency curves and correlation methods introduced by Professor Karl Pearson been very extensively used in direct applications to statistical and biological problems. Of purely statistical writers, we may mention G. Udny Yule, who has published a short treatise entitled "Theory of Statistics" (London, 1911). Numerous excellent memoirs have also appeared in the different English and American mathematical journals and statistical periodicals, especially in the quarterly publication, *Biometrika*, edited by Professor Karl Pearson.

In the above brief sketch, we have only mentioned the most important contributors to the theory of probabilities proper. Numerous able writers have written on the related subject of least squares, the mathematical theory of statistics and insurance mathematics. We shall not discuss the works of these investigators at the present stage. Each of the most important works in the above mentioned branches will receive a short review in the corresponding chapters on statistics and assurance mathematics. The readers interested in the historical development of the theory of probabilities are advised to consult the special treatises on this subject by Todhunter and Czuber.[1]

[1] After this chapter had gone to press I notice that a treatise by the eminent English scholar, Mr. Keynes, is being prepared by The Macmillan Co. In this connection I wish also to call attention to the recent publication by Bachelier (Calcul des probabilites, 1912), a work planned on a broad and extensive scale.—A. F.

CHAPTER III.

THE MATHEMATICAL THEORY OF PROBABILITIES.

12. Definition of Mathematical Probability.—"If our positive knowledge of the effect of a complex of causes is such that we may assume, a priori, t cases as being equally likely to occur, but of which only f, ($f < t$), cases are favorable in causing the event, E, in which we are interested, then we define the proper fraction: $f/t = p$ as the mathematical probability of the happening of the event, E" (Czuber). We might also have defined an a priori probability as the ratio of the equally favorable cases to the co-ordinated possible cases.

As is readily seen, this definition assumes a certain a priori knowledge of the possible and favorable conditions of the event in question, and the probability thus defined is therefore called "a priori probability." Denoting the event by the symbol, E, we express the probability of its occurrence by the symbol $P(E)$, and the probability of its non-occurrence by $P(\bar{E})$. Thus if t is the total number of equally possible cases and f the number of favorable cases for the event, we have:

$$P(E) = \frac{f}{t} = p,$$

and

$$P(\bar{E}) = \frac{t-f}{t} = 1 - \frac{f}{t} = 1 - p = 1 - P(E).$$

This relation evidently gives us: $P(E) + P(\bar{E}) = 1$, which is the symbolic expression for the hypothetical disjunctive judgment that the event E will either happen or not happen. If $f = t$, we have:

$$P(E) = \frac{t}{t} = 1,$$

which is the symbol for the hypothetical judgment that if A exists, E will surely happen. Similarly if $f = 0$, we get

$$P(E) = \frac{0}{t} = 0,$$

or the symbol for the hypothetical judgment: If A exists, E will not happen, or what is the same, $\bar{E}$ will happen.

As we have already mentioned, in an a priori determination of a probability, special stress must be laid upon the requirement that all possible cases must be equally likely to occur. The enumeration of these cases is by no means so easy as may appear at first sight. Even in the most simple problems where there can be doubt about the possible cases being equally likely to occur, it is very easy to make a mistake, and some of the most eminent mathematicians and most acute thinkers have drawn erroneous conclusions in this respect. We shall give a few examples of such errors from the literature on the subject of the theory of probabilities, not on account of their historical interest alone, but also for the benefit of the novice who naturally is exposed to such errors.

13. Example 1.—An Italian nobleman, a professional gambler and an amateur mathematician, had, by continued observation of a game with three dice, noticed that the sum of 10 appeared more often than the sum of 9. He expressed his surprise at this to Galileo and asked for an explanation. The nobleman regarded the following combinations as favorable for the throw of 9:

1	2	6
1	3	5
1	4	4
2	2	5
2	3	4
3	3	3

and for the throw of 10 the six combinations of:

1	3	6
1	4	5
2	2	6
2	3	5
2	4	4
3	3	4

Galileo shows in a treatise entitled "Considerazione sopra il giuco dei dadi" that these combinations cannot be regarded as being equally likely. By painting each of the three dice with the different color it is easy to see that an arrangement such as 1 2 6 can be produced in 6 different ways. Let the colors be white, black and red respectively. We may then make the following arrangements:

White	Black	Red
1	2	6
1	6	2
2	1	6
2	6	1
6	1	2
6	2	1

which gives $3! = 6$ different arrangements. The arrangements of 1 4 4 can be made as follows:

White	Black	Red
1	4	4
4	1	4
4	4	1

which gives 3 different arrangements. The arrangements of 3 3 3 can be made in one way only. By complete enumeration of equally favorable cases we obtain the following scheme:

Sum 9	cases	Sum 10	cases
1, 2, 6	6	1, 3, 6	6
1, 3, 5	6	1, 4, 5	6
1, 4, 4	3	2, 2, 6	3
2, 2, 5	3	2, 3, 5	6
2, 3, 4	6	2, 4, 4	3
3, 3, 3	1	3, 3, 4	3
	25		27

The total number of equally possible cases by the different arrangements of the 18 faces on the dice is $6^3 = 216$. The probability of throwing 9 with three dice is therefore $\frac{25}{216}$, of throwing $10 = \frac{27}{216} = \frac{1}{8}$.

14. Example 2.—D'Alembert, the great French mathematician and natural philosopher and one of the ablest thinkers of his time, assigned $\frac{2}{3}$ as the probability of throwing head at least once in two successive throws with a homogeneous coin. D'Alembert reasons as follows: If head appears first the game is finished and a second throw is not necessary. He therefore gives as equally possible cases (we denote head by H and tail by T): H, TH, TT, and determines thus the probability as $\frac{2}{3}$. Where then is the error of D'Alembert? At first glance the chain of reasoning seems perfect. There are altogether three possible cases of which two are in favor of the event. But are the three cases equally likely? To throw head in a single throw is evidently not the same as to throw head in two successive throws. D'Alembert has left out of consideration the fact that a double throw is allowed. The following analysis shows all the equally possible cases which may occur:

$$HH,\ HT,\ TH,\ TT.$$

Three of those cases favor the event. Hence we have:

$$P(E) = p = \tfrac{3}{4}.$$

We shall return to this problem at a later stage under the discussion of the law of large numbers.

The examples quoted have already shown that the enumeration of the equally likely cases requires a sharp distinction between the different combinations and arrangements of elements. In other words, the solution of the problems requires a knowledge of permutations and combinations. We assume here that the reader is already acquainted with the elements and formulas from the combinatorial analysis and shall therefore proceed with some more illustrations. In the following, when employing the binomial coefficients, we shall use the notation $\binom{m}{k}$ instead of ${}^mC_k$.

15. Example 3.—An urn contains a white and b black balls. A person draws k balls. What is the probability of drawing α white and β black balls?

$$(\alpha + \beta = k, \quad \alpha \leqq a, \quad \beta \leqq b)$$

k balls may be drawn from the urn in as many ways as it is possible to select k elements from $a + b$ elements, which may be done in

$$t = \binom{a+b}{k} = \binom{a+b}{\alpha+\beta}$$

ways. Furthermore there are $\binom{a}{\alpha}$ groups of α white and $\binom{b}{\beta}$ groups of β black balls. Since each combination of any one group of the first groups with any one group of the second groups is favorable for the event, we have as favorable cases:

$$f = \binom{a}{\alpha} \times \binom{b}{\beta}; \quad \text{or} \quad P(E) = \frac{\binom{a}{\alpha} \times \binom{b}{\beta}}{\binom{a+b}{\alpha+\beta}}.$$

Example 4. A special case of the above problem is the following question which often appears in the well known game of whist. What are the respective chances that 0, 1, 2, 3, 4 aces are held by a specified player? There are altogether 52 cards in the game equally distributed among 4 players. Of these cards 4 are aces and 48 are non-aces. Hence we have the following values for a, b, k, α and β.

$a = 4$, $b = 48$, $k = 13$, $\alpha = 0, 1, 2, 3, 4$, $\beta = 13, 12, 11, 10, 9$.

Substituting in the above formula we get:

$$p_0 = \binom{4}{0} \times \binom{48}{13} \div \binom{52}{13} = \frac{82251}{270725}$$

$$p_1 = \binom{4}{1} \times \binom{48}{12} \div \binom{52}{13} = \frac{118807}{270725}$$

$$p_2 = \binom{4}{2} \times \binom{48}{11} \div \binom{52}{13} = \frac{57798}{270725}$$

$$p_3 = \binom{4}{3} \times \binom{48}{10} \div \binom{52}{13} = \frac{11154}{270725}$$

$$p_4 = \binom{4}{4} \times \binom{48}{9} \div \binom{52}{13} = \frac{715}{270725}.$$

A hypothetical disjunctive judgment immediately tells us that in

a game of whist a specified player must either hold 0, 1, 2, 3 or 4 aces. Any such judgment is certain to come true. Hence by adding the 5 above computed probabilities we obtain a check for the accuracy of our calculations. The actual addition of the numerical values of p_0, p_1, p_2, p_3, and p_4 gives us unity which is the mathematical symbol for certainty. Gauss, the renowned German mathematician and astronomer, was an eager whist player. During his forty-eight years of residence in the university town of Göttingen almost every evening he played a rubber of whist with some friends among the university professors. He kept a careful record of the distribution of the aces in each game. After his death these records were found among his papers, headed "Aces in Whist." The actual records agree with the results computed above.

16. Example 5.—An urn contains n similar balls. A part of or all the balls are drawn. What is the probability of drawing an even number of balls?

One ball may be drawn in as many ways as there are balls, two balls in as many ways as we may select two elements out of n elements, and so on. Hence we have for the total number of equally possible cases:

$$t = \binom{n}{1} + \binom{n}{2} + \binom{n}{3} + \cdots + (+1)^n \binom{n}{n}.$$

We have now:

$$(1+1)^n = 1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n},$$

and

$$(1-1)^n = 1 - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^n \binom{n}{n}.$$

The number of favorable cases is given by the expansion:

$$f = \binom{n}{2} + \binom{n}{4} + \cdots.$$

The expression for t is the binominal coefficients less unity. Hence we have:

$$t = (1+1)^n - 1 = 2^n - 1.$$

If we add the two expansions of $(1+1)^n$ and $(1-1)^n$ and then

subtract 2 we get the expansion for $2f$. Hence we have:

$$2f = [(1 + 1)^n + (1 - 1)^n - 2] \quad \therefore f = 2^{n-1} - 1.$$

Thus we shall have as the probability of drawing an even number of balls:

$$p = \frac{2^{n-1} - 1}{2^n - 1}$$

while for an uneven number:

$$q = 1 - p = \frac{2^{n-1}}{2^n - 1}.$$

We notice that the probability of drawing an uneven number of balls is larger than the probability of drawing an even number. This apparently strange result is easily explained without the aid of algebra from the fact that when the urn contains one ball only, we cannot draw an even number. Hence we have $p = 0$, $q = 1$. With two balls we may draw an uneven number in two ways and an even number in one way, thus $p = \frac{1}{3}$, and $q = \frac{2}{3}$. The greater weight of q remains when n is finite; only when

$$n = \infty, \qquad p = q = \tfrac{1}{2}.$$

17. Example 6.—A box contains n balls marked $1, 2, 3, \cdots n$. A person draws n balls in succession and none of the balls thus drawn is put back in the urn. Each drawing is consecutively marked $1, 2, 3, \cdots n$ on n cards. What is the probability that no ball marked α ($\alpha = 1, 2, 3, \cdots n$) appears simultaneously with a drawing card marked α?

The number of equally possible cases is simply the number of permutations of n elements which is equal to $n!$

The number of favorable cases is given by the total number of derangements or relative permutations of n elements, i. e., such permutations wherein the numbers from 1 to n do not appear in their natural places. The formula for such relative permutations was first given by Euler in a memoir of the St. Petersburg Academy entitled "Quaestio Curiosa ex Doctrina Combinationis." Euler makes use of a recursion formula. A German mathematician, Lampe, has, however, derived the formula in a simpler manner in "Grunert's Archives" for 1884.

Lampe denotes by the symbol $\varphi(1)$ the number of permutations wherein 1 does not appear in its natural place. By letting 1 remain fixed in the first place we obtain $(n-1)!$ permutations of the other remaining elements, or:

$$\varphi(1)_n = n! - (n-1)!$$

permutations where 1 is out of place. Of these permutations there are, however, a number wherein 2 appears in its natural place. If we let 2 remain fixed in this place we shall have:

$$\varphi(1)_{n-1} = (n-1)! - (n-2)!$$

permutations wherein 2 is in its place but 1 out of place, there remains thus:

$$\varphi(2)_n = \varphi(1)_n - \varphi(1)_{n-1} = n! - 2(n-1)! + (n-2)!$$

permutations in which neither 1 nor 2 is in its natural place. Letting 3 remain fixed in its place, the remaining $n-1$ elements give:

$$(n-1)! - 2(n-2)! + (n-3)!$$

cases where 3 is in its place but 1 and 2 are not. Accordingly there will be:

$$\varphi(3)_n = \varphi(2)_n - \varphi(2)_{n-1} = n! - 3(n-1)! + 3(n-2)! - (n-3)!$$

permutations in which none of the three elements 1, 2, and 3 is in its place. The complete deduction gives us now for the number r:

$$\varphi(r)_n = n! - \binom{r}{1}(n-1)! + \binom{r}{2}(n-2)! - \cdots + (-1)^r\binom{r}{r}(n-r)!$$

arrangements in which none of the numbers $1, 2, 3, \cdots r$ is in its place. Hence the required probability is:

$$\frac{\varphi(r)_n}{n!} = 1 - \binom{r}{1}\frac{1}{n} + \binom{r}{2}\frac{1}{n(n-1)} - \cdots + (-1)^r\binom{r}{r}\frac{1}{n(n-1)\cdots(n-r+1)};$$

when $n = r$ the above expression becomes:

$$p = 1 - \frac{1}{1} + \frac{1}{1.2} - \frac{1}{1.2.3} + \cdots + (-1)^n \frac{1}{1.2.3 \cdots n}$$

or the probability that none of the balls appear in its numerical order.

When $n = \infty$ the above expression converges towards e^{-1} as a limit. Since the series is rapidly convergent, we may therefore as an approximate value let

$$p = e^{-1} = 0.36788 \cdots.$$

The probability that at least one ball appears in numerical order is

$$q = 1 - p = 0.63213 \cdots.$$

CHAPTER IV.

THE ADDITION AND MULTIPLICATION THEOREMS IN PROBABILITIES.

18. Systematic Treatment by Laplace.—The reader will readily have noticed that the problems hitherto considered have been solved by a direct application of the fundamental definition of a mathematical probability. Almost every branch of pure and applied mathematics has originated in this manner. A few isolated problems, apparently having no mutual connection whatsoever, have presented themselves to different mathematicians. As the number of problems increased, there was found to exist a certain inner relation between them, and from the mere isolated cases there grew a systematic treatment of an entirely new subject.

The theory of probabilities had its origin in games; and the different problems that arose, were treated individually. From the time of Galileo and Cardano to the appearance of Laplace's great treatise, a number of celebrated mathematicians such as Pascal, Fermat, Huyghens, De Moivre, Stirling, Bernoulli and others had solved numerous problems, some of these, as we already have seen in the preceding chapter, of a quite complex nature. But none of these mathematicians had hitherto succeeded in giving a systematic treatment of the subject as a whole. All their treatises were, as any one taking the trouble to look over the works of De Moivre and Bernoulli will readily notice, mere collections of examples solved by direct application of our fundamental definition. It remained for Laplace first to give the definite rules to the science by which the solution of a great number of problems, often very complicated, was reduced to the application of a few stable principles, first given in his "Théorie Analytique des Probabilités" (Paris, 1812).

19. Definition of Technical Terms.—Before entering into a demonstration of Laplace's theorems it will, however, be necessary to explain a few technical terms which seem commonplace

and simple enough but which, nevertheless, must be defined clearly in order to avoid any ambiguity.

In all works on probabilities when speaking of happenings of various events we encounter often the terms, *independent events, dependent events* and *mutually exclusive events*. An event E is said to be independent of another event F when the actual happening of F does not influence in any degree whatsoever the probability of the happening of E. On the other hand, if the probability of E is dependent on or influenced by the previous happening of F, then E is said to be dependent on F. Finally the two events E and F are said to be mutually exclusive when through the occurrence of one of them, say F, the other event E cannot take place, or vice versa. We might also in this case consider the two events E and F as members of a complete disjunction. In a complete hypothetical disjunctive judgment as "When a die is thrown either 1, 2, 3, 4, 5 or 6 will turn up" each member represents a possible event. Any one of these events is mutually exclusive in respect to the other events of the disjunction.

20. The Theorem of the Complete or Total Probability, or the Probability of " Either Or."—When an event, E, may happen in any one of the n different and mutually exclusive ways E_1, E_2, E_3, $\cdots E_n$ with the respective probabilities: $p_1, p_2, p_3, \cdots p_n$, then the probability for the happening of the event, E, is equal to the sum of the individual probabilities: $p_1, p_2, p_3, \cdots p_n$.

Proof: The main event, E, falls in n groups of subsidiary events of which only one can happen in a single trial but of which any one will bring forth the event E. Let us by t denote the total number of equally possible cases. Of these possible cases f are in favor of the event. This favorable group of cases may now be divided into n sub-groups of which f are favorable for the happening of E_1, f_2 in favor of E_2, f_3 in favor of $E_3 \cdots f_n$ in favor of E_n. When we write:

$$P(E) = p = \frac{f}{t} = \frac{f_1 + f_2 + f_3 + \cdots + f_n}{t} = \frac{f_1}{t} + \frac{f_2}{t} + \frac{f_3}{t} + \cdots + \frac{f_n}{t}.$$

Each of the fractions f_a/t ($\alpha = 1, 2, 3, \cdots n$) represents the respective probabilities for the actual occurrence of the subsidiary events, E_1, E_2, E_3, $\cdots E_n$. Hence we shall have

$$P(E) = p = p_1 + p_2 + p_3 + \cdots + p_n.$$

This theorem is also known as the Addition Theorem of probabilities. Instead of "total probability" the German scholar, Reuschle, has suggested the expressive name of the "either or" probability. The term is well selected when we remember that the event, E, will happen when either E_1, or E_2 or $E_3 \cdots$ or E_n happens.

Example 7.—What is the probability to throw 8 with two dice in a single throw?

The total number of ways is $t = 6^2 = 36$. The event in question E is composed of the three subsidiary events favoring the combination of 8:

$$E_1\text{: } 6,\ 2$$
$$E_2\text{: } 5,\ 3$$
$$E_3\text{: } 4,\ 4.$$

Now

$$P(E_1) = \frac{2!}{36} = \frac{1}{18}, \qquad P(E_2) = \frac{2!}{36} = \frac{1}{18}, \qquad P(E_3) = \frac{1}{36}.$$

Hence

$$P(E) = \frac{1}{18} + \frac{1}{18} + \frac{1}{36} = \frac{5}{36}.$$

21. Theorem of the Compound Probability or the Probability of " As Well As."—An event E may happen when every one of the mutually exclusive events E_1, E_2, E_3, $\cdots E_n$ has occurred previously. It is immaterial if the n subsidiary events have happened simultaneously or in succession. But it makes a difference if the events E_1, E_2, E_3, $\cdots E_n$ are independent, or dependent on each other.

1. *Independent Events.*—The probability, $P(E) = p$, for the simultaneous or consecutive appearance of several mutually exclusive events: E_1, E_2, $\cdots E_n$ is equal to the product: $p_1 \cdot p_2 \cdot p_3 \cdot \cdots p_n$ of the individual probabilities of the n events.

Proof: Let the number of possible cases entering into the complex that brings forth the event E be t. Each of the t_1

possible cases corresponding to the event E_1 may occur simultaneously with each one of the t_2 cases corresponding to the event E_2. Thus we have altogether $t_1 \times t_2$ cases falling on E_1 and E_2 at the same time. Continuing in the same way of reasoning it is readily seen that the total number of equally possible cases resulting from the simultaneous occurrence of the events E_1, E_2, E_3, $\cdots E_n$ is equal to $t_1 \times t_2 \times t_3 \times \cdots t_n$. By applying the same reasoning to the favorable cases we get as their total number:

$$f = f_1 \times f_2 \times f_3 \times \cdots f_n.$$

Hence the final probability for the happening of the simultaneous or consecutive appearance of the n minor events is:

$$P(E) = \frac{f}{t} = \frac{f_1}{t_1} \times \frac{f_2}{t_2} \times \frac{f_3}{t_3} \times \cdots \frac{f_n}{t_n} = p_1 \times p_2 \times p_3 \times \cdots p_n.$$

Example 8.—A card is drawn from a whist deck, another card is drawn from a pinochle deck. What is the probability that they both are aces?

A whist deck contains 52 cards of which four are aces, a pinochle deck 48 cards with 8 aces. Denoting the probabilities of getting an ace from the whist and pinochle decks by $P(E_1)$ and $P(E_2)$ respectively we have:

$$P(E) = P(E_1)P(E_2) = \frac{4}{52} \times \frac{8}{48} = \frac{1}{78}.$$

2. *Dependent Events.*—The n events E_1, E_2, E_3, $\cdots E_n$ are not independent of each other, but are related in such a way that the appearance of E_1 influences E_2, that event influences in turn E_3, E_3 event E_4 and so on.

The same reason holds as above, and,

$$P(E) = p = p_1 \times p_2 \times p_3 \times \cdots p_n.$$

But p_2 means here the probability for the happening of E_2 after the actual occurrence of E_1, p_3 the probability for the happening of E_3 after E_1 and E_2 have previously happened, and so on for all n events.

Example 9.—A card is drawn from a whist deck and replaced by a joker, and then a second card is drawn. What is the probability that both cards are aces?

Denoting the two subsidiary events by E_1 and E_2 we have:

$$P(E) = P(E_1)P(E_2) = \frac{4}{52}\cdot\frac{3}{52} = \frac{3}{13 \times 52} = \frac{3}{676}.$$

The two above theorems are known as the multiplication theorems in probabilities. Reuschle has also suggested the name "the as well as probability."

22. Poincaré's Proof of the Addition and Multiplication Theorem.—The French mathematician and physicist, H. Poincaré, has derived the above theorems in a new and elegant manner in his excellent little treatise: "Lecons sur le Calcul des Probabilités," Paris, 1896.

Poincaré's proof is briefly as follows:

Let E_1 and E_2 be two arbitrary events.

E_1 and E_2 may happen in α different ways.

E_1 may happen but not E_2 in β different ways.

E_2 may happen but not E_1 in γ different ways.

Neither E_1 nor E_2 will happen in δ different ways.

We assume the total $\alpha + \beta + \gamma + \delta$ cases to be equally likely to occur.

The probability for the occurrence of E_1 is

$$p_1 = \frac{\alpha + \beta}{\alpha + \beta + \gamma + \delta}.$$

The probability for the occurrence of E_2 is

$$p_2 = \frac{\alpha + \gamma}{\alpha + \beta + \gamma + \delta}.$$

The probability for the occurrence of at least one of the events E_1 and E_2 is

$$p_3 = \frac{\alpha + \beta + \gamma}{\alpha + \beta + \gamma + \delta}.$$

The probability for the occurrence of both E_1 and E_2 is

$$p_4 = \frac{\alpha}{\alpha + \beta + \gamma + \delta}.$$

The probability for the occurrence of E_1 when E_2 has already occurred is

$$p_5 = \frac{\alpha}{\alpha + \gamma}.$$

The probability for the occurrence of E_2 when E_1 has already occurred is

$$p_6 = \frac{\alpha}{\alpha + \beta}.$$

The probability for the occurrence of E_1 when E_2 has not already occurred is

$$p_7 = \frac{\beta}{\beta + \gamma}.$$

The probability for the occurrence of E_2 when E_1 has not already occurred is

$$p_8 = \frac{\gamma}{\gamma + \delta}.$$

We have now the following identical relations:

$$p_1 + p_2 = p_3 + p_4, \qquad p_3 = p_1 + p_2 - p_4,$$

i. e., the probability that of two arbitrary events at least one will happen is equal to the probability that the first will happen plus the probability that the second will happen less the probability that both will happen. The particular problem which we may happen to investigate may possibly be of such a nature that the two events E_1 and E_2 cannot happen at the same time, in that case $p_4 = 0$, and we get:

$$p_3 = p_1 + p_4.$$

In this equation we immediately recognize the addition theorem for two mutually exclusive events. By substitution of the proper values we have furthermore:

$$p_4 = p_2 \cdot p_5 \quad \text{or} \quad p_4 = p_1 \cdot p_6.$$

These equations contain the theorems proved under § 21, of the probability for two mutually dependent events.

23. Relative Probabilities.—We shall now finally give an alternative demonstration of the same two theorems. It will, of course, be of benefit to the student to see the subject from as many view points as possible; moreover, the following remarks will contain some very useful hints for the solution of more complicated problems by the application of so-called " relative prob-

abilities "and a few elementary theorems from the calculus of logic. The following paragraphs are mainly based upon a treatise in the Proceedings of the Royal Academy of Saxony, by the German mathematician and actuary, F. Hausdorff.

In our fundamental definition of a mathematical probability for the happening of an event E, expressed in symbols by $P(E)$, as the ratio of the equally favorable and equally possible cases resulting from a general complex of causes, we were able to compute the so-called ordinary or absolute probabilities. But if we, from among the favorable cases and possible cases, select only such as bring forward a certain different event, say F, then we obtain the "relative probability" for the happening of E under the assumption that the subsidiary event, F, has occurred previously. For this relative probability we shall employ the symbol $P_F(E)$, which reads "the relative probability of E, positi F." The following problem illustrates the meaning of relative probabilities. If an honor card is drawn from an ordinary deck of cards, what is the probability that it is a king? Denoting the subsidiary event of drawing an honor card by F, and the main event of drawing a king by E, we may write the above mentioned probability in the symbolic form: $P_F(E)$. If on the other hand we knew a priori that a king was drawn, we may also ask for the probability of having drawn an honor card. Since any king also is an honor card, we may write in symbols: $P_E(F) = 1$.

Before entering upon the immediate determination of relative probabilities we shall first define a few symbols from the calculus of logic. We denote first of all the occurrence of an event E by E, the non-occurrence of the same event by $\overline{E}$. Similarly we have for the occurrence and non-occurrence of other events, $F, G, H, \cdots$ and $\overline{F}, \overline{G}, \overline{H}, \cdots$. $E + F$ means that at least one of the two events E and F will happen. $E \times F$ or simply $E \cdot F$ means the occurrence of both E and F. From the above definition it follows immediately that $\overline{E + F} = \overline{E} \cdot \overline{F}$ and $E = E \cdot F + E \cdot \overline{F}$.

This last relation simply states that E will happen when either E and F happen simultaneously or when E and the non-appearance of F happen at the same time. If furthermore F_1, $\overline{F}_1$, F_2,

$\bar{F}_2 \cdots F_n, \bar{F}_n$ constitute the members of a complete disjunction, i. e., mutually exclusive events, we have in general:

$$E = E \cdot F_1 + E \cdot \bar{F}_1 + E \cdot F_2 + E \cdot \bar{F}_2 + \cdots E \cdot F_n + E \cdot \bar{F}_n.$$

From the original definition of a probability, it follows now:

$$P(E) = P(E \cdot F) + P(E \cdot \bar{F}),$$

and

$$P(E) = P(E \cdot F_1) + P(E \cdot \bar{F}_1) + P(E \cdot F_2) + P(E \cdot \bar{F}_2) + P(E \cdot F_n) + P(E \cdot \bar{F}_n),$$

i. e., the probability that of several mutually exclusive events one at least will happen is the sum of the probabilities of the happening of the separate events. This is the symbolic form for the addition theorem.

24. Multiplication Theorem.—We next take two arbitrary events. From these events we may form the following combinations:

$E \cdot F,\ E \cdot \bar{F},\ \bar{E} \cdot F,\ \bar{E} \cdot \bar{F}$, i. e.,
Both E and F happen,
E happens but not F
F happens but not E
Neither E nor F happens.

Furthermore let $\alpha, \beta, \gamma, \delta$, be the respective numbers of the favorable cases for the above four combinations of the events E and F. Following the previous method of Poincaré, we shall have:

$$P(E) = \frac{\alpha + \beta}{\alpha + \beta + \gamma + \delta}, \qquad P(F) = \frac{\alpha + \gamma}{\alpha + \beta + \gamma + \delta},$$

$$P_F(E) = \frac{\alpha}{\alpha + \gamma}, \qquad P_E(F) = \frac{\alpha}{\alpha + \beta},$$

$$P(E \cdot F) = \frac{\alpha}{\alpha + \beta + \gamma + \delta}.$$

25. Probability of Repetitions.—From the above equations it immediately follows:

$$P(E \cdot F) = P(E) \times P_E(F) = P(F) \times P_F(E),$$

which is the symbolic form for the multiplication theorems of compound probabilities.

In special cases it may happen that the different subsidiary events: E_1, E_2, $E_3 \cdots E_n$ are all similar. We shall then have, following the symbolic method:

$$E = E_1 \cdot E_2 \cdot E_3 \cdots E_n = E_1 \cdot E_1 \cdot E_1 \cdots E_1 = E_1^n,$$

and

$$P(E) = P(E_1^n) = P(E_1)^n.$$

This gives us the following theorem:

The probability for the repetition n times of a certain event, E, is equal to the nth power of its absolute probability.

Thus if $P(E) = p$ we have immediately $P(\bar{E}) = 1 - p$.

$$P(E^n) = P(E)^n = p^n.$$
$$P(\bar{E}^n) = P(\bar{E})^n = (1 - p)^n.$$

Thus the probability for the occurrence of E at least once in n trials is

$$P(E + E + \cdots n \text{ times}) = 1 - P(\bar{E}^n) = 1 - (1 - p)^n.$$

Denoting the numerical quantity of this probability by Q we have:

$$1 - Q = (1 - p)^n.$$

Solving this equation for n we shall have:

$$n = \frac{\log (1 - Q)}{\log (1 - p)}.$$

Whenever n equals, or is greater than, the above logarithmic value for given values of Q and p we are sure that Q will exceed a previously given proper fraction. To illustrate:

Example 10.—How often must a die be thrown so that the probability that a six appears at least once is greater than $\frac{1}{2}$?

Here $p = \frac{1}{6}$, $Q = \frac{1}{2}$. Hence we must select for n the smallest positive integer satisfying the relation:

$$\frac{\log (1 - \frac{1}{2})}{\log (1 - \frac{1}{6})} = \frac{\log \frac{1}{2}}{\log \frac{5}{6}} = \frac{.301035}{.079186}\text{; i. e., } n = 4.$$

For this particular value of n we have in reality:

$$Q = 1 - (\tfrac{5}{6})^4 = .518.$$

26. Application of the Addition and Multiplication Theorems in Problems in Probabilities.—We shall next proceed to illustrate the theorems of the preceding paragraphs by a few examples. First, we shall apply the demonstrated theorems to some of the examples we have already solved by a direct application of the fundamental definition of a mathematical probability.

Example 11.—We take first of all our old friend, the problem of D'Alembert. What is the probability of throwing head at least once in two successive throws with an uniform coin?

This problem is most easily solved by finding the probability first for not getting head in two successive throws. By the multiplication theorem this probability is: $p = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Then the probability to get head at least once is $1 - \frac{1}{4} = \frac{3}{4}$ from a simple application of the rule in § 25. A more lengthy analysis is as follows. Denoting the event by E, the following cases may appear which may bring forth the desired event: Head in first throw which we shall denote by H_1 and head in second throw which we denote by H_2, or head in first throw (H_1) and tail in second (T_2), or finally tail in first (T_1) and head in second (H_2). Then we have:

$$E = H_1 \cdot H_2 + H_1 \cdot T_2 + T_1 \cdot H_2,$$

or:

$$P(E) = P(H_1) \cdot P(H_2) + P(H_1) \cdot P(T_2) + P(T_1) \cdot P(H_2)$$
$$= \tfrac{1}{2} \times \tfrac{1}{2} + \tfrac{1}{2} \times \tfrac{1}{2} + \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{3}{4}.$$

27. Example 12.—What is the probability of throwing at least twelve in a single throw with three dice? The expected event occurs when either 12, 13, 14, . . . or 18 is thrown. Of these events only one may happen at a time. We may, therefore, apply the addition theorem and obtain as the total probability:

$$p = p_{12} + p_{13} + p_{14} + \cdots + p_{18}.$$

where $p_{12}, p_{13}, \cdots p_{18}$ are the respective probabilities for throwing the sums of 12, 13, $\cdots$ or 18. These subsidiary probabilities were determined in § 13 under the problem of Galileo, and:

$$p = \tfrac{25}{216} + \tfrac{21}{216} + \tfrac{15}{216} + \tfrac{10}{216} + \tfrac{6}{216} + \tfrac{3}{216} + \tfrac{1}{216} = \tfrac{27}{72}.$$

28. Example 13.—An urn contains a white, b black and c red balls. A single ball is drawn $\alpha + \beta + \gamma$ times in succession, and the ball thus drawn is replaced before the next drawing takes place. To determine the probability that (1) there are first α white, then β black and finally γ red balls, (2) the drawn balls appear in three closed groups of α white, β black and γ red balls, but the order of these groups is arbitrary, (3) that white, black and red balls appear in the same number as above, but in any order whatsoever.

1. Denoting the three subsidiary events for drawing α white, β black and γ red balls by F_1, F_2 and F_3, and the main event for drawing the balls in the prescribed order by E, we may write the probability for the occurrence of the main event in following symbolic form involving symbolic probabilities:

$$P(E) = P(F_1)P_{F_1}(F_2)P_{F_1F_2}(F_3).$$

Substituting the algebraic values for $P(F_1)$, $P(F_2)$ and $P(F_3)$ in the expression for $P(E)$, and then applying Hausdorff's rule (§ 24) we get:

$$P(E) = p_1 = \frac{a^\alpha}{(a+b+c)^\alpha} \times \frac{b^\beta}{(a+b+c)^\beta} \times \frac{c^\gamma}{(a+b+c)^\gamma}$$

$$= \frac{a^\alpha b^\beta c^\gamma}{(a+b+c)^{\alpha+\beta+\gamma}}.$$

2. In the second part of the problem the order of the three different groups is immaterial. The three subsidiary events: F_1, F_2 and F_3, may therefore be arranged in any order whatsoever. The total number of arrangements is $3! = 6$. The probability of the happening of any one of these arrangements separately is the same as the probability computed under (1). By applying the addition theorem we get therefore as the probability of the occurrence of this event:

$$p_2 = \frac{6a^\alpha b^\beta c^\gamma}{(a+b+c)^{\alpha+\beta+\gamma}}.$$

3. The third part is more easily solved by a direct application of the definition of a mathematical probability. The order of the balls drawn is here immaterial. Of each individual com-

bination of α white, β black balls and γ red balls it is possible to form $(a + \beta + \gamma)!/\alpha!\beta!\gamma!$ different permutations as the total number of favorable cases. The above number of equally possible cases is here $(a + b + c)^{\alpha+\beta+\gamma}$. Hence we have:

$$p_3 = \frac{(\alpha + \beta + \gamma)!}{\alpha!\beta!\gamma!} \times \frac{a^\alpha b^\beta c^\gamma}{(a + b + c)^{\alpha+\beta+\gamma}}.$$

29. **Example 14.**—In an urn are n balls among which are α white and β black. What is the probability in three successive drawings to draw (1) first two white and then one black ball, (2) two white and one black ball in any order whatsoever? $(\alpha+\beta\leqq n)$. The probability to draw first one white, then another white and finally a black ball is:

$$p_1 = \frac{\alpha}{n}\frac{(\alpha - 1)}{(n - 1)} \times \frac{\beta}{(n - 2)}.$$

The probability for any of the other arrangements is the same, or we have for (2)

$$p_2 = 3p_1 = \frac{3\alpha}{n}\frac{(\alpha - 1)}{(n - 1)} \times \frac{\beta}{(n - 2)}.$$

30. **Example 15.**—What is the chance to throw a doublet of 6 at least once in n consecutive throws with two dice? (Pascal's Problem.)

Chevalier de Mere, a French nobleman and a great friend of all games of chance, went more deeply into the complex of causes in different games than most of the ordinary gamblers of his time. Although not a proficient mathematician he understood sufficient, nevertheless, to give some very interesting problems for which he got the ideas from the gambling resorts he frequented. De Mere was a friend of the great French mathematician and philosopher, Blaise Pascal, and went to him whenever he wanted information on some apparently obscure point in the different games in which he participated. The chevalier had from patient observation noticed that he could profitably bet to throw a six at least once in four throws with a single die. He reasoned now that the number of throws to throw a doublet at least once with two dice ought to be proportional to the corresponding equal number of possible cases with a single die. For one die there are

6 possible cases, for two 36. Thus de Mere thought he could safely bet to throw a doublet of 6 in 24 throws with two dice. An actual trial by several games of dice proved extremely disastrous to the finances of the nobleman, who then went to Pascal for an explanation. Pascal solved the problem by a direct application of the definition of a mathematical probability. We shall, however, solve it by an application of the multiplication theorem.

The probability to get a doublet of 6 in a single throw is $\frac{1}{36}$. The probability of not getting a double six is therefore $1 - \frac{1}{36} = \frac{35}{36}$. The probability of the happening of this event n consecutive times is $(\frac{35}{36})^n$. Thus the probability of getting a double six at least once in n throws with two dice becomes: $p = 1 - (\frac{35}{36})^n$. Solving this equation for n we shall have:

$$n = \frac{\log (1 - p)}{\log 35 - \log 36},$$

for $p = \frac{1}{2}$ we shall have:

$$n = \frac{\log 2}{\log 36 - \log 35} = 24.6 \cdots.$$

First for 25 throws we may bet safely one to one while for 24 throws such a bet was unfavorable. This shows the fallacy of de Mere's reasoning.

31. Example 16.—An urn, A, contains a balls of which α are white, another similar urn, B, contains b balls of which β are white. A single ball is drawn from one of the two urns. What is the probability that the ball is white? The beginner may easily make the following error in the solution of this problem. The probability to get a white ball from A is α/a, from B, β/b. Thus the total probability to get a white ball is: $\alpha/a + \beta/b$. This result is, however, wrong, for we may, by selecting proper values for a, b, α and β, obtain a total probability which in numerical value is greater than unity. Thus if $a = 7$, $b = 7$, $\alpha = 5$, $\beta = 4$, we get as the total probability:

$$p = \tfrac{5}{7} + \tfrac{4}{7} = \tfrac{9}{7}.$$

This result is evidently wrong, since a mathematical probability is never an improper fraction. The error lies in the fact that we

have regarded the two events of drawing a ball from either urn as independent and mutually exclusive. A simple application of the symbolic rule for relative probabilities will give us the result immediately. The main event, E, is composed of the two following subsidiary events: (1) to get a white ball from A, or (2) to get a white ball from B. We shall symbolically denote these two events by $A \cdot W$ and $B \cdot W$ respectively. Thus we have:

$$P(E) = P(A \cdot W) + P(B \cdot W) = P(A)P_A(W) + P(B)P_B(W).$$

Now the probability to obtain urn A is $P(A) = p_1 = \frac{1}{2}$, also to get B: $P(B) = p_2 = \frac{1}{2}$. The probability to get a white ball from A when this particular urn is previously selected is expressed by the relative probability:

$$P_A(W) = p_3 = \frac{\alpha}{a}.$$

Similarly for B:

$$P_B(W) = p_4 = \frac{\beta}{b}.$$

Substituting these different values in the expression for $P(E)$ we get finally:

$$P(E) = p = \frac{1}{2} \times \frac{\alpha}{a} + \frac{1}{2} \times \frac{\beta}{b} = \frac{1}{2}\left(\frac{\alpha}{a} + \frac{\beta}{b}\right).$$

For the particular numerical example we have:

$$p = \frac{1}{2}\left(\frac{5}{7} + \frac{4}{7}\right) = \frac{9}{14}.$$

32. Example 17.—The probability of the happening of a certain event, E, is p, while the probability for the non-occurrence of the same event is $q = 1 - p$. The trial is now to be repeated n times. The probability that there will be first α successes and then β failures is:

$$P(E^\alpha)P_{E^\alpha}(\overline{E}^\beta) = p^\alpha \cdot q^\beta (\alpha + \beta = n).$$

This is the probability that the two complementary events E and $\overline{E}$ happen in the order prescribed above. When the order, in which the successes and failures happen, plays no rôle during the n trials, that is to say it is only required to obtain α successes

and β failures in any order whatsoever in n total trials, then the arrangement of the α factors p and β factors q is immaterial. The total number of arrangements of n elements of which α are equal to p and β equal to q is simply $n!/(\alpha! \times \beta!)$. For any one particular arrangement of α factors p and β factors q the probability of the happening of the two complementary events in this particular arrangement is equal to $p^\alpha \cdot q^\beta$. The Addition Theorem immediately gives the answer for α successes and β failures in any order whatsoever as:

$$P(\dot{E}^\alpha \cdot \bar{E}^\beta) = p_\alpha = \binom{n}{\alpha} p^\alpha q^{n-\alpha}.$$

Let us, for the present, regard this probability as being a function of the variable quantity, α, (n being a constant quantity). We may then write:

$$p_\alpha = \varphi(\alpha).$$

Letting α assume all positive integral values from 0 to n the above expression for p_α becomes:

$$p_0 = \binom{n}{0} p^0 \cdot q^n, \quad p_1 = \binom{n}{1} p \cdot q^{n-1},$$

$$p_2 = \binom{n}{2} p^2 \cdot q^{n-2}, \quad \cdots \quad p_n = p^n.$$

These are the respective probabilities for no successes, one success, two successes, . . . and finally n successes in n trials. The above quantities are, however, merely the different members of the binomial expansion $(p + q)^n$. Since $p + q = 1$ from the nature of the problem, we also have $(p + q)^n = 1$, or $p_0 + p_1 + p_2 + \cdots + p_n = 1$. This last equation is the symbolic form for the simple hypothetical disjunctive judgment: E must happen either 0, 1, 2, $\cdots$ or n times in n total trials. We shall return to this problem later under the discussion of the Bernoullian Theorem. In fact, the above example constitutes an essential part of this famous theorem which has proven one of the most important and far reaching in the whole theory of probability.

33. Example 18. De Moivre's Problem.—The following problem was first given by the eminent French-English mathemati-

cian, Abraham de Moivre, in a treatise, entitled "De Mensura Sortis," which was published in London about 1711.

An urn contains $n + 1$ balls marked $0, 1, 2, \cdots n$. A person makes i drawings in succession, and each ball is put back in the urn before the next drawing takes place. What is the probability that the sum of the numbers on the n balls thus drawn equals s?

The first ball may be drawn in $n + 1$ ways, the second ball may also be drawn in $n + 1$ ways. Hence two balls may be drawn in $(n + 1)^2$ ways or i balls in $(n + 1)^i$ ways: This is the total number of equally possible cases.

If we expand the expression:

$$(x^0 + x^1 + x^2 + x^3 + x^4 + \cdots x^n)^i \tag{1}$$

after the multinominal theorem, we notice that the coefficient to x^s arises out of the different ways in which $0, 1, 2, 3, \cdots n$ can be grouped together so as to form s by addition, which also is the total number of favorable cases. The expression (1) inside the bracket represents a geometrical progression, which may be written as:

$$(1 - x^{n+1})^i(1 - x)^{-i} = \left\{1 - ix^{n+1} + \binom{i}{2}x^{2n+2} - \binom{i}{3}x^{3n+3} + \cdots\right\} \times \left\{1 + ix + \binom{i+1}{2}x^2 + \binom{i+2}{3}x^3 + \cdots\right\}.$$

By actual multiplication we get a power series in x. The terms containing x^s are obtained in the following manner: the first term of the first factor being multiplied with the term

$$\binom{i + s - 1}{s} x^s \text{ of the second factor,}$$

the second term of the first factor multiplied with the term:

$$\binom{i + s - n - 2}{s - n - 1} x^{s-n-1} \text{ of the second factor,}$$

the third term of the first factor multiplied with the term:

$$\binom{i + s - 2n - 3}{s - 2n - 2} x^{s-2n-2} \text{ of the second factor.}$$

. . . .

Thus the coefficient of x^s is equal to

$$\binom{i+s-1}{s} - \binom{i}{1}\binom{i+s-n-2}{s-n-1} + \binom{i}{2}\binom{i+s-2n-3}{s-2n-2} - \cdots.$$

The above expression may by further reductions be brought to the form:

$$\frac{(s+1)(s+2)\cdots(s+i-1)}{1\cdot 2\cdots(i-1)} - \binom{i}{1}\frac{(s-n)(s-n+1)\cdots(s-n+i-2)}{1\cdot 2\cdots(i-1)} + \binom{i}{2}\frac{(s-2n-1)(s-2n)\cdots(s-2n+i-3)}{1\cdot 2\cdots(i-1)} - \cdots.$$

The series breaks of course as soon as negative factors appear in the numerator. The required probability is therefore

$$P_s = \frac{1}{(n+1)^i}\left\{\frac{(s+1)(s+2)\cdots(s+i-1)}{1\cdot 2\cdots(i-1)} - \binom{i}{1}\frac{(s-n)(s-n+1)\cdots(s-n+i-2)}{1\cdot 2\cdots(i-1)} + \cdots\right\}.$$

34. **Example 19.**—If a single experiment or observation is made on n pairs of opposite (complementary) events, E_α and $\bar{E}_\alpha$ with the respective probabilities of happening p_α and q_α ($\alpha = 1, 2, 3, \cdots n$), to determine the probability that: (1) exactly r, (2) at least r of the events E_α will happen.

This problem is of great importance, especially in life assurance mathematics. It happens frequently that an actuary is called upon to determine the probability that exactly r persons will be alive m years from now out of a group of n persons of any age whatsoever, each person's age and his individual coefficient of survival through the period being known beforehand.

Various demonstrations have been given of this problem. The first elementary proof was probably due to Mr. George King,

the English actuary, in his well-known text-book. The Austrian mathematician and actuary, E. Czuber, has simplified King's method in his "Wahrscheinlichkeitsrechnung" (1903). Later the Italian actuary, Toja, has given an elegant proof in *Bolletino degli Attuari*, Vol. 12. Finally another Italian mathematician, P. Medolaghi, has investigated the problem from the standpoint of symbolic logic. In the following we shall adhere to the demonstration of Czuber and also give a short outline of the symbolic method.

In order to answer the first part of the problem we must form all possible combinations of r factors of p and $n - r$ factors of q and then sum all such combinations of n factors. Denoting the event by $E_{[r]}$ we have:

$$\begin{aligned} P(E_{[r]}) &= \Sigma p_\alpha p_\beta p_\gamma \cdots q_\lambda q_\epsilon \cdots q_\omega \\ &= \Sigma p_\alpha p_\beta \cdots (1 - p_\lambda)(1 - p_\epsilon) \cdots (1 - p_\omega). \end{aligned} \tag{1}$$

We shall now denote the *sum* of all products in (1) containing φ factors p by the symbol S_ϕ. It is readily seen that φ will have all positive integral values from r to n inclusive. We may therefore write the total compound probability in the following form:

$$P(E_{[r]}) = A_0 S_r + A_1 S_{r+1} + A_2 S_{r+2} + \cdots + A_{n-r} S_n. \tag{2}$$

The student must bear in mind that the different S are merely symbols for different *sums* of all the products of r, $r + 1$, $r + 2$, $\cdots n$ factors p respectively. Our problem is now to determine the unknown coefficients A. It is easily seen that the coefficient $A_0 = 1$, since all different products containing r factors p appear only once. The other coefficients of the form A do not depend on the values of p, however. They remain therefore unaltered if we equate all of the various p's and let them equal p. Expression (1) then simply becomes $\binom{n}{r} \cdot p^r(1 - p)^{n-r}$. We must form all possible rth powers of n similar factors, which can be done in $\binom{n}{r}$ ways. The expression (2) on the other hand

becomes:

$$\binom{n}{r} \cdot p^r + A_1\binom{n}{r+1} \cdot p^{r+1} + A_2\binom{n}{r+2} \cdot p^{r+2} + \cdots + A_{n-r}p^n.$$

Any S_ϕ is by definition the sum of all products containing φ factors p and we may form $\binom{n}{\varphi}$ such products from n elements p. But we saw above that φ might only have all positive values from r to n inclusive, hence expression (2) will naturally take the above form. We have therefore

$$\binom{n}{r} \cdot p^r(1-p)^{n-r} = \binom{n}{r} \cdot p^r + A_1\binom{n}{r+1} p^{r+1} + A_2\binom{n}{r+2} \cdot p^{n+2} + \cdots + A_{n-r}p^n.$$

Expanding the expression on the left hand side by means of the binomial theorem and equating the coefficients of equal powers of p, we get:

$$-\binom{n}{r}\binom{n-r}{1} = A_1\binom{n}{r+1}; \quad \binom{n}{r}\binom{n-r}{2} = A_2\binom{n}{r+2}$$

$$-\binom{n}{r}\binom{n-r}{3} = A_3\binom{n}{r+3}, \quad \cdots \quad (-1)^{n-r}\binom{n}{r}\binom{n-r}{n-r} = A_{n-r},$$

or:

$$A_1 = -\binom{r+1}{1}, \quad A_2 = \binom{r+2}{2}, \quad A_3 = -\binom{r+3}{3}, \quad \cdots$$

$$A_{n-r} = (-1)^{n-r}\binom{n}{n-r}.$$

Substituting these values in (2) for the unknown coefficients A, we shall have:

$$P(E_{[r]}) = S_r - \binom{r+1}{1} S_{r+1} + \binom{r+2}{2} S_{r+2} - \cdots + (-1)^{n-r}\binom{n}{n-r} S_n.$$

If we expand the algebraic expression:

$$\frac{S^r}{(1+S)^{r+1}} = S^r(1+S)^{-(r+1)}$$

we have:

$$S^r - \binom{r+1}{1}S^{r+1} + \binom{r+2}{2}S^{r+2} - \cdots\cdots$$
$$+ (-1)^{n-r}\binom{n}{n-r}S^n - \cdots.$$

We may therefore write $P(E) = \dfrac{S^r}{(1+S)^{r+1}}$, when every exponent is replaced by an index number (*i. e.*, S^ϕ replaced by S_ϕ) and the expansion broken off at the term S^n. The student must of course constantly bear in mind the symbolic meaning of S_ϕ.

The second part of the problem is easily solved by the symbolic method. Denoting this particular event by E_r, we have the following identity:

$$P(E_r) - P(E_{r+1}) = P(E_{[r]})$$

or

$$P(E_r) - P(E_{[r]}) = P(E_{r+1}).$$

The following relations are self-evident:

$P(E_0) = 1;$

$P(E_{[0]}) = \dfrac{S^0}{1+S} = \dfrac{1}{1+S};$

$P(E_1) = P(E_0) - P(E_{[0]}) = 1 - \dfrac{1}{1+S}$, also;

$P(E_2) = P(E_1) - P(E_{[1]}) = \dfrac{S}{1+S} - \dfrac{S}{(1+S)^2} = \dfrac{S^2}{(1+S)^2}.$

The complete induction gives us finally:

$$P(E_r) = \frac{S^r}{(1+S)^r}.$$

Assuming the rule is true for r, we may easily prove it is true for $r+1$ also. We have in fact:

$$\frac{S^r}{(1+S)^r} - \frac{S^r}{(1+S)^{r+1}} = \frac{S^{r+1}}{(1+S)^{r+1}}.$$

35. Example 20. Tchebycheff's Problem.—The following solution of a very interesting problem is due to the eminent Russian mathematician, Tchebycheff, one of the foremost of modern analysts.

A proper fraction is chosen at random. What is the probability it is in its lowest terms?

Stated in a slightly different wording the same question may also be put as follows: If A/B is a proper fraction, what is the probability that A and B are prime to each other?

If $p_2, p_3, p_5, \cdots p_m$ denote respectively the probabilities that each of the primes $2, 3, 5, \cdots m$ *is not* a common factor of numerator and denominator of A/B, then the probability that *no* prime number is a common factor is:

$$P = p_2 \cdot p_3 \cdot p_5 \cdots p_m \cdots p_s \cdots \text{ad. inf.} \qquad \text{(I)}$$

This follows from the multiplication theorem and from the fact that the sequence of prime numbers is infinite.

Tchebycheff now first finds the probability $q_m = 1 - p_m$ that the fraction A/B *does* contain the prime m as factor of both A and B. By dividing any integral number by the prime m we obtain besides the quotient a certain remainder that must be one of the following numbers, viz.:

$$0, 1, 2, 3, 4, \cdots (m - 1).$$

Each of the above remainders may be regarded as a possible event. The probability to obtain 0 as a remainder is accordingly $1/m$. The probability that m is contained as a factor of A is therefore $1/m$. This same quantity is also the probability that m is a factor of B. The probability that both A and B are divisible by m is therefore:

$$q_m = 1 - p_m = \frac{1}{m} \cdot \frac{1}{m} = \frac{1}{m^2}, \quad \text{or} \quad p_m = 1 - \frac{1}{m^2}.$$

Hence we have for the various primes:

$$p_2 = 1 - \frac{1}{2^2}, \qquad p_3 = 1 - \frac{1}{3^2}, \qquad p_5 = 1 - \frac{1}{5^2}, \quad \cdots.$$

Formula (I) then takes the form:

$$P = \left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right)\left(1 - \frac{1}{5^2}\right) \cdots \text{ad. inf.} \qquad \text{(II)}$$

Forming the reciprocal $1/P$ we get:

$$\frac{1}{P} = \frac{1}{1 - \frac{1}{2^2}} \cdot \frac{1}{1 - \frac{1}{3^2}} \cdot \frac{1}{1 - \frac{1}{5^2}} \cdots \text{ad. inf.}$$

Now each factor on the right hand side is the sum of a geometrical progression, as:

$$\frac{1}{P} = \left(1 + \frac{1}{2^2} + \frac{1}{(2^2)^2} + \cdots\right)\left(1 + \frac{1}{3^2} + \frac{1}{(3^2)^2} + \cdots\right)$$
$$\left(1 + \frac{1}{5^2} + \frac{1}{(5^2)^2} + \cdots\right) \cdots \text{ad. inf.}$$

Multiplying out we shall have:

$$\frac{1}{P} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \cdots \text{ad. inf.}$$

The above infinite series is, however, merely the well known Eulerian expression for $\pi^2/6$, hence:

$$P = \frac{6}{\pi^2}.$$

Suppose furthermore we were assured that none of the three primes 2, 3, 5 was a common factor of both A and B. What would then be the probability that the fraction might be reduced by division by one or more of the other primes?

Denoting by the symbol $P_{(7)}$ the probability that *none* of the primes from 7 and upwards is a common factor, we get:

$$P_{(7)} = \left(1 - \frac{1}{7^2}\right)\left(1 - \frac{1}{11^2}\right)\left(1 - \frac{1}{13^2}\right) \cdots \text{ad. inf.},$$

also:

$$P = \frac{6}{\pi^2} = \left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right)\left(1 - \frac{1}{5^2}\right) P_{(7)},$$

or:

$$P_{(7)} = \frac{6}{\pi^2} \div \left[\left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right)\left(1 - \frac{1}{5^2}\right)\right] = 0.950.$$

The probability of the divisibility of both numerator and denominator of a fraction chosen at random by a prime larger than 5 is thus:

$$1 - P_{(7)} = \frac{1}{20}.$$

The summation of the infinite series of the reciprocals of the squares of the natural numbers baffled for a long time the skill of some of the most eminent mathematicians. Jacob Bernoulli, the renowned classical writer on probabilities, proved its convergency but failed to find its sum. The final summation was first performed by Euler.

CHAPTER V.

MATHEMATICAL EXPECTATION.

36. Definition, Mean Values.—It is common belief among many people that gambling and all kinds of betting have their source in reckless desire. This is often argued by moral reformers, but cannot be said to be the true cause. Whenever by ordinary gambling or by a bet, actual value is exposed to a complete or partial loss, this exposure is not due to the fact that the gamester is reckless, but because there is hope of an actual gain. "Hope," says Spinoza in his treatise on ethics, "is the indeterminate joy caused by the conception of a future state of affairs of whose outcome we are in doubt." Actual mathematical calculation cannot be attempted on the basis of this definition any more than it could be attempted to determine a mathematical probability from the definition of Aristotle. "We disregard therefore the psychological element of desire, which is associated with hope or expectation as well as the anxiousness or dread associated with the related psychological element of non-desire" (Cantor).

The so-called *mathematical expectation* is the product of an expected gain in actual value and the mathematical probability of obtaining such a gain. The danger of loss may in this case be regarded as a negative gain. Thus if a person, A, may expect the gain, G, from the event, E, whose probability of happening is equal to p, then $e = p \cdot G$ is his mathematical expectation. The quantity expressed by the symbol, e, is here the amount it is safe to hazard for the expected gain, G. We may also regard the quantity, e, as a mean value or average value. Among a large number of n cases only np will bring the gain, G, the others not. Thus the total gain is:

$$pnG \div n = pG.$$

Suppose we have n mutually exclusive events, $E_1, E_2, \cdots, E_n$,

forming a complete disjunction. For their respective probabilities we have then the following equation:

$$p_1 + p_2 + p_3 + \cdots + p_n = 1.$$

If the actual occurrence of a certain one of these events, say, E_α, brings a gain of G_α, then the total value of the mathematical expectation of the n events is:

$$e = p_1 \cdot G_1 + p_2 \cdot G_2 + \cdots + p_n \cdot G_n = \Sigma p_\alpha \cdot G_\alpha.$$

Since $\Sigma p_\alpha = 1$ this result may be written:

$$e \times (p_1 + p_2 + \cdots + p_n) = G_1 \cdot p_1 + G_2 \cdot p_2 + G_3 \cdot p_3 + \cdots G_n \cdot p_n,$$

hence e may be regarded as the mean value of the different quantities G_α with the weights p_α ($\alpha = 1, 2, 3, \cdots, n$).

Although we shall discuss the theory of mean values in a following chapter a few preliminary remarks might not be out of place here.

A variable quantity X is related to a series of events E_1, E_2, E_3, $\cdots$, E_n (it being assumed that these events form a complete disjunction) in such a manner that when E_α happens X takes on the value x_α ($\alpha = 1, 2, 3, \cdots, n$). If furthermore $p_1, p_2, p_3, \cdots$ denote the respective probabilities of the occurrence of E_1, E_2, E_3, $\cdots$, then

$$M(X) = p_1x_1 + p_2x_2 + \cdots p_nx_n$$

is called the mean value or simply the mean of X.

The above definition may be illustrated by the following concrete urn-scheme. An urn contains N balls of which a_1 balls are marked x_1, a_2 balls marked x_2 $\cdots$ and finally a_n balls marked x_n where $a_1 + a_2 + a_3 + \cdots a_n = N$. Each drawing from the urn produces a certain number X, which may assume n different values $x_1, x_2, x_3, \cdots, x_n$, each with the respective probabilities:

$$p_1 = \frac{a_1}{N},\ p_2 = \frac{a_2}{N} \cdots p_n = \frac{a_n}{N}.$$

The arithmetic mean of all the numbers written on the balls is:

$$\frac{a_1x_1 + a_2x_2 + \cdots a_nx_n}{N},$$

which agrees with the mean as defined above.

37. The Petrograd (St. Petersburg) Problem.—In this connection it is worthy to note a celebrated problem, which on account of its paradoxical nature has become a veritable stumbling block, and has been discussed by some of the most eminent writers on probabilities. The problem was first suggested by Daniel Bernoulli in a communication to the Petrograd—or as it was then called St. Petersburger Academy—in 1738.

The Petrograd problem may shortly be stated as follows: Two persons A and B are interested in a game of tossing a coin under the following conditions. An ordinary coin is tossed until head turns up, which is the deciding event. If head turns up the first time A pays one dollar to B, if head appears first at the second toss B is to receive two dollars, if first at the third time four dollars and so on. What is the mathematical expectation of B? Or in other words, how much must B pay to A before the game starts in order that the game may be considered fair?

The mathematical expectation of B in the first trial is $\frac{1}{2} \times 1 = \frac{1}{2}$. The mathematical expectation for head in second throw is $(\frac{1}{2})^2 \times 2 = \frac{1}{2}$. Or in general the mathematical probability that head appears for the first time in the nth toss is $(\frac{1}{2})^n$, and the co-ordinated expectation is $2^{n-1} \div 2^n = \frac{1}{2}$. Thus the total expectation is expressed by the following series:

$$\tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \cdots.$$

When $n = \infty$ as its limiting value it thus appears that B could afford to pay an infinite amount of money for his expected gain.

38. Various Explanations of the Paradox. The Moral Expectation.—This evidently paradoxical result has called forth a number of explanations of various forms by some eminent mathematicians. One of the commentators was D'Alembert. It was to be expected that the famous encyclopædist, who — as we have seen — did not view the theory of probabilities in too kindly a manner, would not hesitate to attack. He returns repeatedly to this problem in the "Opuscules" (1761) and in "Doutes et questions" (Amsterdam, 1770).

D'Alembert distinguishes between two forms of possibilities, viz., metaphysical and physical possibilities. An event is by

him called a metaphysical possibility, when it is not absurd. When the event is not too "uncommon" in the ordinary course of happenings it is a physical possibility. That head would appear for the first time after 1,000 throws is metaphysically possible but quite impossible physically. This contention is rather bold. "What would," as Czuber remarks, "D'Alembert have said to an actual reported case in 'Grunert's Archiv' where in a game of whist each of the four players held 13 cards of one suit." The numerical probability of such an event as expressed by mathematical probabilities is $(635013559600)^{-4}$.

D'Alembert's definitions including the half metaphorical term "ordinary course" are rather vague. And what numerical value of the mathematical probability constitutes the physical impossibility? D'Alembert gives three arbitrary solutions for the probability of getting head in the nth throw, namely:

$$\frac{1}{2^n(1+\beta n^2)},\quad \frac{1}{2^{n+\alpha n}},\quad \frac{1}{2^n+\dfrac{2^nB}{(K-n)^{q/2}}},$$

where α, β, B, K are constants and q an uneven number.

Daniel Bernoulli himself gives a solution wherein he introduces the term "moral expectation." If a person possesses a sum of money equal to x then according to Bernoulli

$$dy = \frac{kdx}{x}$$

is the moral expectation of x, k being a constant quantity. Integrating we get:

$$\int_a^b dy = k\int_a^b \frac{dx}{x} = k(\log b - \log a) = k\log\frac{b}{a},$$

which is the moral expectation of an increase $b - a$ of an original value a. If now x denotes the sum owned by B we may replace the mathematical expectation by their corresponding moral expectations, that is to say replace $2^{n-1}/2^n$ by $(1/2^n)\log((\alpha+2^{n-1})/x)$ and we then have:

$$k\left(\tfrac{1}{2}\log\frac{x+1}{x}+\tfrac{1}{4}\log\frac{x+2}{x}+\cdots\frac{1}{2^n}\log\frac{x+2^n}{x}\right),$$

which is a convergent series. In this connection, it may be mentioned that the Bernoullian hypothesis has found quite an extensive use in the modern theory of utility.

De Morgan in his splendid little treatise "On Probabilities" takes the view that the solution as first given is by no means an anomaly. He quotes an actual experiment in coin tossing by Buffon. Out of 2,048 trials 1,061 gave head at the first toss, 494 at the second, 232 at the third, 137 at the fourth, 56 at the fifth, 29 at the sixth, 25 at the seventh, 8 at the eighth and 6 at the ninth. Computing the various mathematical expectations, we find that the maximum value is found in the 25 sets with head in the seventh toss, which gives a gain of $25 \times 64 = 1{,}600$. The most rare occurrence, the 6 sets of head in the ninth throw gives a gain of $6 \times 256 = 1{,}536$, which is the next highest gain in all the nine sets. De Morgan furthermore contends that if Buffon had tried a thousand times as many games, the results would not only have given more, but *more per game*, arguing "that a larger net would have caught not only more fish but more varieties of fish; and in two millions of sets, we might have seen cases in which head did not appear till the twentieth throw." Furthermore, "the player might continue until he had realized not only any given sum, but any given sum per game." Therefore according to De Morgan the mathematical expectation of a player in a single game must be infinite.

CHAPTER VI.

PROBABILITY A POSTERIORI.

39. Bayes's Rule. A Posteriori Probabilities.—The problems hitherto considered have all had certain points in common. Before entering upon the calculations of the mathematical probability of the happening of the event in question, we knew beforehand a certain complex of causes which operated in the general domain of action. We also were able to separate this general complex of productive causes into two distinctive minor domains of complexes, of which one would bring forth the event, E, while the other domain would act towards the production of the opposite event, $\bar{E}$. Furthermore we also were able to measure the respective quantitative magnitudes of the two domains, and then, by a simple algebraic operation, determine the probability as a proper fraction. The addition and multiplication theorems did not introduce any new principles, but only gave us a set of systematic rules which facilitated and shortened the calculations of the relations between the different absolute probabilities. The above method of determination of a mathematical probability is known as an a priori determination, and such probabilities are termed a priori probabilities.

The problems treated in the preceding chapters have, nearly all, been related to different games of chance, or purely abstract mathematical quantities. The inorganic nature of this kind of problems has made it possible for us to treat them in a relatively simple manner. In many of the problems, which we shall consider hereafter, organic elements enter as a dominant factor and make the analysis much more complicated and difficult.

All social and biological investigations, which are of a much larger benefit and practical value than the problems in games of chance, lead often to a completely different category of probability problems, which are known as "a posteriori probabilities." In problems where organic life enters into the calculations, the complex of productive causes is so varied and manifold, that

our minds are not able to pigeonhole the different productive causes, placing them in their proper domains of action. But we know that such causes do exist and are the origin of the event. If now, by a series of observations, we have noticed the actual occurrence of the event, E (or the occurrence of the opposite event $\bar{E}$), the problem of the determination of an a posteriori probability to find the probability that the event E originated from a certain complex, say F. We must then, first of all, form a complete hypothetical judgment of the form: E either happens from the complexes F_1, or F_2, or F_3, $\cdots$ or F_n. But we must not forget that, in general, the different complexes F_α ($\alpha = 1, 2, \cdots, n$) of the disjunction are not known a priori. We must, therefore, determine the respective probabilities for the actual existence of such disjunctive complexes F_α. These probabilities of existence for the complexes of causes are in general different for each member, a fact which often has been overlooked by many investigators and writers on a posteriori probabilities, and which has given rise to meaningless and paradoxical results.

40. Discovery and History of the Rule.—The first discoverer of the rule for the computation of a posteriori probabilities by a purely deductive process was the English clergyman, T. Bayes. Bayes's treatise was first published after the death of the author by his friend, Dr. Price, in *Philosophical Transactions* for 1763. The treatise by the English clergyman was, for a long time, almost forgotten, even by the author's own countrymen; and later English writers have lost sight of the true "Bayes's Rule" and substituted a false, or to be more accurate, a special case of the exact rule, in the different algebraic texts, under the discussion of the so called "inverse probabilities," a name which is due probably to de Morgan, and which in itself is a great misnomer. This point, presently, we shall discuss in detail.

The careless application of the exact rule has recently led to a certain distrust of the whole theory of "a posteriori probabilities." Scandinavian mathematicians were probably the first to criticize the theory. In 1879, Mr. J. Bing, a Danish actuary, took a very critical attitude towards the mathematical principles underlying Bayes's Rule, in a scholarly article in the mathe-

matical journal *Tidsskrift for Matematik.* Bing's article caused a sharp, and often heated, discussion among the older and younger Danish mathematicians at that time; but his views seem to have gained the upper hand, and even so great an authority on the whole subject as the late Dr. T. Thiele, in his well-known work, "Theory of Observations" (London, 1903), refers to Bing's article as "a crushing proof of the fallacies underlying the determination of a posteriori probabilities by a purely deductive method." As recently as 1908, the Danish writer on philosophy, Dr. Kroman, has taken up cudgels in defense of Bayes in a contribution in the *Transactions of the Royal Danish Academy of Science,* which has done much towards the removal of many obscure and erroneous views of the older authors. Among English writers, Professor Chrystal, in a lecture delivered before the Actuarial Society of Edinburgh, has also given a sharp criticism of the rule, although he does not go so deeply into the real nature of the problem as either Bing or Kroman.

Despite Chrystal's advice to "bury the laws of inverse probabilities decently out of sight, and not embalm them in text books and examination papers" the old view still holds sway in recent professional examination papers. It is therefore absolutely necessary for the student preparing for professional examinations to be acquainted with the theory. In the following paragraphs we shall, therefore, give the mathematical theory of Bayes's Rule with several examples illustrating its application to actual problems, together with a criticism of the rule.

41. Bayes's Rule (Case I).—(*The different complexes of causes producing the observed event, E, possess different a priori probabilities of existence.*) Let E denote a certain state or condition, which can appear under only one of the mutually exclusive complexes of causes: F_1, F_2, $\cdots$ and not otherwise. Let the probability for the actual existence of F_1 be κ_1 and if F_1 really exists then let ω_1 be the "productive probability" for bringing forth the observed event, E (E being of a different nature from F), which can only occur after the previous existence of one of the mutually exclusive complexes, F. Let, in the same manner, F_2 have an "existence probability" of κ_2 and a "productive probability" of ω_2, F_3 an existence probability of κ_3 and a pro-

ductive probability of $\omega_3 \cdots$ etc. If now, by actual observation, we have noted that the event E has occurred exactly m times in n trials, then the probability that the complex F_1 was the origin of E is:

$$Q_1 = \frac{\kappa_1 \cdot \omega_1^m(1 - \omega_1)^{n-m}}{\Sigma\kappa_a \cdot \omega_a^m(1 - \omega_a)^{n-m}} \quad (\alpha = 1, 2, 3, \cdots).$$

Similarly that complex F_2 was the origin:

$$Q_2 = \frac{\kappa_2 \cdot \omega_2^m(1 - \omega_2)^{n-m}}{\Sigma\kappa_a \cdot \omega_a^m(1 - \omega_a)^{n-m}}$$

and so on for the other complexes.

Proof.—Let the number of equally possible cases in the general domain of action, which leads to one of the complexes F_a, be t. Furthermore, of these t cases let f_1 be favorable for the existence of complex F_1, f_2 for F_2, f_3 for $F_3, \cdots$, etc. Then the probabilities for the existence of the different complexes F_a ($\alpha = 1, 2, 3, \cdots n$) are:

$$\kappa_1 = \frac{f_1}{t}, \qquad \kappa_2 = \frac{f_2}{t}, \qquad \kappa_3 = \frac{f_3}{t} \cdots \text{ respectively.}$$

Of the f_1 favorable cases for complex F_1, λ_1 are also favorable for the occurrence of E.

Of the f_2 favorable cases for complex F_2, λ_2 are also favorable for the occurrence of E.

Of the f_3 favorable cases for complex F_3, λ_3 are also favorable for the occurrence of E.

. . . .

The probability of the happening of E under the assumption that F_1 exists, i. e., the relative probability: $P_{F_1}(E)$, is:

$$\omega_1 = \frac{\lambda_1}{f_1}$$

or in general:

$$\omega_a = \frac{\lambda_a}{f_a} \qquad (\alpha = 1, 2, 3, \cdots).$$

The total number of equally likely cases for the simultaneous occurrence of the event E with either one of the favorable cases

for $F_1, F_2, F_3, \cdots$ is:

$$\lambda_1 + \lambda_2 + \lambda_3 + \cdots = \Sigma\lambda_\alpha.$$

The number of favorable cases for the simultaneous occurrence of F_1 and E is λ_1, for the simultaneous occurrence of F_2 and E, $\lambda_2, \cdots$, etc. Hence: we have as measures for their corresponding probabilities

$$Q_1 = \frac{\lambda_1}{\Sigma\lambda_\alpha}, \qquad Q_2 = \frac{\lambda_2}{\Sigma\lambda_\alpha}, \qquad \cdots.$$

But

$$\lambda_1 = \omega_1 \cdot f_1, \quad \lambda_2 = \omega_2 \cdot f_2, \quad \cdots, \quad \text{etc.},$$

and

$$f_1 = \kappa_1 \cdot t, \quad f_2 = \kappa_2 \cdot t, \quad \cdots, \quad \text{etc.}$$

Hence

$$\lambda_1 = \omega_1 \cdot \kappa_1 \cdot t, \quad \lambda_2 = \omega_2 \cdot \kappa_2 \cdot t, \quad \cdots, \text{etc.}$$

Substituting these values in the above expression for $Q_1, Q_2, \cdots$ we get:

$$Q_1 = \frac{\kappa_1 \cdot \omega_1}{\Sigma k_\alpha \cdot \omega_\alpha}, \qquad Q_2 = \frac{\kappa_2 \cdot \omega_2}{\Sigma\kappa_\alpha \cdot \omega_\alpha}, \qquad \cdots$$

as the respective probabilities that the observed event originated from the complexes $F_1, F_2, F_3, \cdots$. Such probabilities are called a posteriori probabilities.

Let us now for a moment investigate the above expression for $Q_1, Q_2, \cdots$. The numerator in the expression for Q_1 is $\kappa_1 \cdot \omega_1$. But κ_1 is simply the a priori probability for the existence of F_1 while ω_1 is the a priori productive probability of bringing forth the event observed from complex F_1. The product $\kappa_1 \cdot \omega_1$ is simply the relative probability $P_{F_1}(E)$, or the probability that the event E originated from F_1. In the denominator we have the expression $\Sigma\kappa_\alpha\omega_\alpha$ ($\alpha = 1, 2, \cdots n$) which is the total probability to get E from any of the complexes F_α. From example 17 (Chapter IV) we know that the probability to get E exactly m times from F_1 in n total trials is:

$$p_1 = \binom{n}{m} \kappa_1 \cdot \omega_1{}^m(1 - \omega_1)^{n-m}$$

and the probability to get E from any one of the complexes, F,

m times out of n is:

$$\Sigma p_a = \binom{n}{m} \Sigma \kappa_a \cdot \omega_a^m (1 - \omega_a)^{n-m} \quad (\alpha = 1, 2, 3, \cdots).$$

If, by actual observation, we know the event E to have happened exactly m times out of n, then the a posteriori probability that F_1 was the origin is:

$$Q_1 = \frac{\binom{n}{m} \kappa_1 \cdot \omega_1^m (1 - \omega_1)^{n-m}}{\Sigma \binom{n}{m} \kappa_a \cdot \omega_a^m (1 - \omega_a)^{n-m}} \quad (\alpha = 1, 2, 3, \cdots). \quad \text{(I)}$$

The factorials $\binom{m}{n}$ in numerator and denominator cancel each other of course. It will be noticed that, in the above proof, it is not assumed that the a posteriori probability is proportional to the a priori probability, an assumption usually made in the ordinary texts on algebra.

42. Bayes's Rule (Case II).—(*Special Case. The a priori probabilities of existence of the different complexes are equal.*) Sometimes the different complexes F may be of such special characters that their a priori probabilities of existence are equal, i. e.,

$$\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 \cdots \kappa_n.$$

In this case the equation (I) simply reduces to:

$$Q_1 = \frac{\omega_1^m (1 - \omega_1)^{n-m}}{\Sigma \omega_a^m (1 - \omega_a)^{n-m}}. \quad \text{(II)}$$

Equation (I) gives, however, the most general expression for Bayes's Rule which may be stated as follows:

If a definite observed event, E, can originate from a certain series of mutually exclusive complexes, F, and if the actual occurrence of the event has been observed, then the probability that it originated from a specified complex or a specified group of complexes is also the "a posteriori" probability or probability of existence of the specified complex or group of complexes.

43. Determination of the Probabilities of Future Events Based Upon Actual Observations.—It happens frequently that

our knowledge of the general domain of action is so incomplete, that we are not able to determine, a priori, the probability of the occurrence of a certain expected event. As we already have stated in the introduction to a posteriori probabilities, this is nearly always the case with problems wherein organic life enters as a determining factor or momentum. But the same state of affairs may also occur in the category of problems relating to games of chance, which we have hitherto considered. Suppose we had an urn which was known to contain white and black balls only, but the actual ratio in which the balls of the two different colors were mixed, was unknown. With this knowledge beforehand, we should not be able to determine the probability for the drawing of a white ball. If, on the other hand, we knew, from actual experience by repeated observations, the results of former drawings from the same urn when the conditions in the general domain of action remained unchanged during each separate drawing, then these results might be used in the determination of the probability of a specified event by future drawings.

Our problem may be stated in its most general form as follows: Let F_a denote a certain state or condition in the general domain of action, which state or condition can appear only in one or the other of the mutually exclusive forms: $F_1, F_2, F_3, \cdots$, and not otherwise. Let the probability of existence of $F_1, F_2, F_3, \cdots$ be $\kappa_1, \kappa_2, \kappa_3, \cdots$ respectively, and when one of the complexes F_1, F_2, $F_3, \cdots$ exists (occurs) let $\omega_1, \omega_2, \omega_3, \cdots$ be the respective productive probabilities of bringing forth a specified event, E. If now, by actual observation, we know the event, E, to have happened exactly m times out of n total trials (the conditions in the general domain of action being the same at each individual trial), what is then the probability that the event, E, will happen in the $(n+1)$th trial also?

By Bayes's Rule we determined the "a posteriori" probabilities or the probabilities of existence of the complexes $F_1, F_2, \cdots$ as:

$$Q_1 = \frac{\kappa_1 \cdot \omega_1{}^m(1-\omega_1)^{n-m}}{\Sigma\kappa_a \cdot \omega_a{}^m(1-\omega_a)^{n-m}}, \quad Q_2 = \frac{\kappa_2 \cdot \omega_2{}^m(1-\omega_2)^{n-m}}{\Sigma\kappa_a \cdot \omega_a{}^m(1-\omega_a)^{n-m}}, \quad \cdots$$
$$(\alpha = 1, 2, 3, \cdots).$$

In the $(n+1)$th trial E may happen from any one of the mutually

exclusive complexes: $F_1, F_2, F_3, \cdots$ whose respective probabilities in producing the event, E, are $\omega_1, \omega_2, \omega_3, \cdots$. The addition theorem then gives us as the total probability of the occurrence of E in the $(n + 1)$th trial:

$$R_a = \Sigma P_{F_a}(E) = Q_1 \cdot \omega_1 + Q_2 \cdot \omega_2 + Q_3 \cdot \omega_3 + \cdots = \frac{\Sigma \kappa_a \cdot \omega_a^m (1 - \omega_a)^{n-m} \cdot \omega_a}{\Sigma \kappa_a \cdot \omega_a^m (1 - \omega_a)^{n-m}} \quad (\alpha = 1, 2, 3, \cdots). \tag{III}$$

If the a priori probabilities of existence are of equal magnitude (Case II) the factors κ in the above expression cancel each other in numerator and denominator and we have

$$R = \frac{\Sigma \omega_a^m (1 - \omega_a)^{n-m} \omega_a}{\Sigma \omega_a^m (1 - \omega_a)^{n-m}}. \tag{IV}$$

44. Examples on the Application of Bayes's Rule.—*Example 21.*—An urn contains two balls, white or black or both kinds. What is the probability of getting a white ball in the first drawing, and if this event has happened and the ball replaced, what is then the probability to get white in the following drawing?

Three conditions are here possible in the urn. There may be 0, 1, or 2 white balls. Each hypothetical condition has a probability of existence equal to $\frac{1}{3}$, and the productive probabilities for white are 0, $\frac{1}{2}$ and 1 respectively. The total probability to get white is therefore:

$$p_1 = \tfrac{1}{3} \cdot 0 + \tfrac{1}{3} \cdot \tfrac{1}{2} + \tfrac{1}{3} \cdot 1 = \tfrac{1}{2}.$$

If we now draw a white ball then the probabilities that it came from the complexes: F_1, F_2, F_3, respectively, are:

$$0 \div \tfrac{1}{2}, \quad \tfrac{1}{6} \div \tfrac{1}{2}, \quad \tfrac{1}{3} \div \tfrac{1}{2}.$$

These are also new existence probabilities of the three probabilities. The probability for white in second drawing is therefore.

$$(0 \div \tfrac{1}{2})0 + (\tfrac{1}{6} \div \tfrac{1}{2})\tfrac{1}{2} + (\tfrac{1}{3} \div \tfrac{1}{2})1 = \tfrac{5}{6}.$$

This solution of the problem is, however, not a unique solution, because it is an arbitrary solution. It is arbitrary in this respect, that we have without further consideration given all three complexes the same probability of existence, $\frac{1}{3}$. We shall discuss

this part of the question under the chapter on the criticism of Bayes's Rule.

Example 22.—An urn contains five balls of which a part is known to be white and the rest black. A ball is drawn four times in succession and replaced after each drawing. By three of such drawings a white ball was obtained and by one drawing a black ball. What is the probability that we will get a white ball in the fifth drawing?

In regard to the contents of the urn the following four hypotheses are possible:

F_1:	4	white,	1	black	balls,
F_2:	3	"	2	"	"
F_3:	2	"	3	"	"
F_4:	1	"	4	"	"

Since we do not know anything about the ratio of distribution of the different colored balls, we may by a direct application of the principle of insufficient reason regard the four complexes as equally probable, or:

$$\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = \tfrac{1}{4}.$$

If either F_1, F_2, F_3 or F_4 exists, the respective productive probabilities are:

$$\omega_1 = \tfrac{4}{5}, \quad \omega_2 = \tfrac{3}{5}, \quad \omega_3 = \tfrac{2}{5}, \quad \omega_4 = \tfrac{1}{5}.$$

By a direct substitution in the formula:

$$R = \frac{\Sigma \omega_a{}^m (1 - \omega_a)^{n-m} \cdot \omega_a}{\Sigma \omega_a{}^m (1 - \omega_a)^{n-m}}$$

($\alpha = 1, 2, 3, 4$) for $n = 4$ and $m = 3$ we get:

$$R = \frac{(\frac{4}{5})^3(\frac{1}{5})(\frac{4}{5}) + (\frac{3}{5})^3(\frac{2}{5})(\frac{3}{5}) + (\frac{2}{5})^3(\frac{3}{5})(\frac{2}{5}) + (\frac{1}{5})^3(\frac{4}{5})(\frac{1}{5})}{(\frac{4}{5})^3(\frac{1}{5}) + (\frac{3}{5})^3(\frac{2}{5}) + (\frac{2}{5})^3(\frac{3}{5}) + (\frac{1}{5})^3(\frac{4}{5})} = \tfrac{47}{73}.$$

45. Criticism of Bayes's Rule.—In most English treatises on the theory of chance the "a posteriori" determination of a mathematical probability is discussed under the socalled "inverse probabilities." This somewhat misleading name was probably first introduced by the eminent English mathematician and actuary, Augustus de Morgan. In the opening of the discussion

of a posteriori probabilities in the third chapter of his treatise, "An Essay on Probabilities," de Morgan says: "In the preceding chapter, we have calculated the chances of an event, knowing the circumstances under which it is to happen or fail. We are now to place ourselves in an inverted position, we know the event, and ask what is the probability which results from the event in favor of any set of circumstances under which the same might have happened." Is this now an inverse process? By the a priori or—as de Morgan prefers to call them,—the direct probabilities, we started from a definitely known condition and determined the probability for a future event, E, or what is the same, the probability of a specified future state of affairs. Here we start knowing the present condition and try to determine a past condition. The process apparently appears to be the inverse of the former, although they both are the same. We possess a definite knowledge of a certain condition and try to determine the probability of the existence of a specified state of affairs, in general different from the first condition, but whether this state of affairs occurred in the past or is to occur in the future has no bearing on our problem. In other words, time does not enter as a determining factor. And even if we were willing to admit the two processes of the determination of the different probabilities to be inverse, the probabilities themselves can not be said to be inverse. Nevertheless, this misleading name appears over and over again in examination papers in England and in America as a thoroughly embalmed corpse which ought to have been buried long ago. What is really needed, is a change of customary nomenclature in the whole theory of probability. Instead of direct and inverse, a priori and a posteriori probabilities, it would be more proper to speak about "prospective" and "retrospective" probabilities in the application of Bayes's Rule. All probabilities are in reality determined by an empirical process. That there is a certain probability to throw a six with a die we only know after we have formed a definite conception of a die. The only probabilities which we perhaps rightly may name a priori are the arbitrary probabilities in purely mathematical problems where we assume an ideal state of affairs. "There is," to quote the Danish writer on logic, Dr. Kroman, "really

more reason to doubt the a priori than the a posteriori probabilities, and it would be more natural and also more exact in the application of Bayes's Rule to speak about the actual or original and the new or gained probability."

The discussion above has really no direct bearing on Bayes's Rule but was introduced in order to give the student a clearer understanding of the main principles underlying the whole determination of a posteriori probabilities by means of actual experimental observations, and also to remove some obscure points. From his ordinary mathematical training every student of mathematics has an almost intuitive understanding of an inverse process. Naturally when he encounters again and again the customary heading: " inverse probabilities " in text-books he obtains from the very start—almost before he starts to read this particular chapter—an inverse idea of the subject instead of the idea he really ought to have. Nowhere in continental texts on the theory of probabilities, will the reader be able to find the words direct and inverse applied in the same sense as in English texts since the introduction of these terms by de Morgan. We shall advise readers who have become accustomed to the old terms to pay no serious attention to them.

46. Theory Versus Practice.—In § 41 we reduced Bayes's Rule to its most general form:

$$Q = \frac{\kappa_a \cdot \omega_a{}^m(1 - \omega_a)^{n-m}}{\Sigma\kappa_a \cdot \omega_a{}^m(1 - \omega_a)^{n-m}} \quad (\alpha = 1, 2, 3, \cdots).$$

This is an exact expression for the rule, but it is at the same time almost impossible to employ it in practice. Only in a few exceptional cases do we know, a priori, the different values of the often numerous probabilities of existence κ_a, of the complexes F_a, and in order to apply the rule with exact results we require here sufficient facts and information about the different complexes of causes from which the observed event, E, originated. Bayes deduced the rule from special examples resulting from drawings of balls of different colors from an urn where the different complexes of causes were materially existent. The probability of a cause or a certain complex of causes did not here mean the probability of existence of such a complex but the probability

that the observed event originated from this particular complex. In order to elucidate this statement we give following simple example:

Example 23.—A bag contains 4 coins, of which one is coined with two heads, the other three having both head and tail. A coin is drawn at random and tossed four times in succession and each time head turns up. What is the probability that it was the coin with two heads?

The two complexes F_1 and F_2, which may produce the event, E, are: F_1, the coin with two heads, and F_2, an ordinary coin. The probability of existence of F_1 is the probability of drawing the single coin with two heads which is equal to $\frac{1}{4}$, the probability of existence for the other complex, F_2, is equal to $\frac{3}{4}$. The respective productive probabilities are 1 and $\frac{1}{2}$. Thus $\kappa_1 = \frac{1}{4}$, $\kappa_2 = \frac{3}{4}$, $\omega_1 = 1$ and $\omega_2 = \frac{1}{2}$. Substituting these values in formula (I) $(n = 4,\ m = 4)$, we get:

$$Q = (\tfrac{1}{4} \times 1^4) \div (\tfrac{1}{4} \times 1^4 + \tfrac{3}{4} \times (\tfrac{1}{2})^4) = \tfrac{1}{4} \div \tfrac{19}{64} = \tfrac{16}{19}.$$

But in most cases we do not know anything about the material existence of the complexes of causes from which the event, E, originated. On the contrary, we are forced to form a hypothesis about their actual existence. To start with a simple case we take example 21 of § 44.

We assumed here three equally possible conditions in the urn before the drawings, namely the presence of 0, 1, or 2 white balls. From this assumption we found the probability to get a white ball in the second drawing, after we had previously drawn a white ball and then put it back in the urn before the second drawing, to be equal to $\frac{5}{6}$. As we already remarked, this solution is not unique because it is an arbitrary solution. It is arbitrary to assign, without any consideration whatsoever, $\frac{1}{3}$ as the probability of existence to each of the three conditions. Let us suppose that each of the two balls bore the numbers 1 and 2 respectively. We may then form the following equally likely conditions:

$$b_1b_2,\ b_1w_2,\ b_2w_1,\ w_1w_2,$$

each condition having an a priori probability of existence equal to $\frac{1}{4}$ and a productive probability for the drawing of a white

ball equal to: 0, $\frac{1}{2}$, $\frac{1}{2}$ and 1 respectively. Thus:

$$\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = \tfrac{1}{4}$$

and

$$\omega_1 = 0, \quad \omega_2 = \tfrac{1}{2}, \quad \omega_3 = \tfrac{1}{2}, \quad \omega_4 = 1.$$

The respective a posteriori probabilities, that is the new or gained probabilities of the four hypothetical conditions, become now by the application of Bayes's Rule (Formula II):

$$Q_1 = \tfrac{0}{2}, \quad Q_2 = \tfrac{1}{2} \div 2, \quad Q_3 = \tfrac{1}{2} \div 2, \quad Q_4 = \tfrac{1}{2}.$$

Hence the probability for white in the second drawing is:

$$\left(\text{Formula IV: } R = \frac{\Sigma\omega_a{}^m(1 - \omega_a)^{n-m}\omega_a}{\Sigma\omega_a{}^m(1 - \omega_a)^{n-m}}\right)$$

$$R = 0 \div 2 + (\tfrac{1}{4} \div 2) + (\tfrac{1}{4} \div 2) + (1 \div 2) = \tfrac{3}{4}.$$

In the first solution we got $\frac{5}{6}$ for the same probability. Which answer is now the true one? Neither one! The true answer to the problem is that it is not given in such a form that the last question—the probability of getting a white ball in the second drawing—may be settled without any doubt. The answer must be conditional. Following the first hypothesis we got $\frac{5}{6}$, while the second hypothesis gives $\frac{3}{4}$ as the answer.

We next proceed to example 22 which is almost identical in form to the first one, the only difference being a greater variety of hypothetical conditions. We started here with the following four hypotheses:

F_1: 4 white, 1 black ball, F_2: 3 white, 2 black, F_3: 2 white, 3 black and F_4: 1 white and 4 black balls, assigning $\frac{1}{4}$ as the hypothetical existence probability.

By marking the 5 balls similarly as in the last example, with the numbers from 1 to 5 we may form the complexes:

F_1: 4 white and 1 black ball in $\binom{5}{1}$ ways,
F_2: 3 " " 2 " balls " $\binom{5}{2}$ "
F_3: 2 " " 3 " " " $\binom{5}{2}$ "
F_4=1 " " 4 " " " $\binom{5}{1}$ "

This gives us a total of $5 + 10 + 10 + 5 = 30$ different complexes. Assuming all of these complexes equally likely

to occur, we get following probabilities of existence and productive probabilities:

$\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = \cdots = \kappa_{30} = \frac{1}{30}$

$\omega_1 = \omega_2 = \omega_3 = \omega_4 = \omega_5 = \frac{4}{5}$ (Productive prob. for F_1)

$\omega_6 = \omega_7 = \omega_8 = \cdots = \omega_{15} = \frac{3}{5}$ (Productive prob. for F_2)

$\omega_{16} = \omega_{17} = \cdots = \omega_{25} = \frac{2}{5}$ (Productive prob. for F_3)

$\omega_{26} = \omega_{27} = \omega_{28} = \omega_{29} = \omega_{30} = \frac{1}{5}$ (Productive prob. for F_4).

The total probability of getting a white ball in the second drawing is now $\frac{\Sigma\omega_\alpha^3(1-\omega_\alpha)\omega_\alpha}{\Sigma\omega_\alpha^3(1-\omega_\alpha)}$ $(\alpha = 1, 2, 3, \cdots, 30)$.

Actual substitution of the above values of ω in this formula gives us the final result as: $R = \frac{17}{28}$.

47. Probabilities Expressed by Integrals.—By making an extended use of the infinitesimal calculus Mr. Bing and Dr. Kroman in their memoirs arrived at much more ambiguous results through an application of the rule of Bayes. Starting with the fundamental rule as given in equation (I) in § 41, we may at times encounter somewhat simpler conditions inside the domain of causes. The total complex of actions may embrace a large number of smaller sub-complexes construed in such a way that the change from one complex to another may be regarded as a continuous process, so that the productive probabilities are increased by an infinitely small quantity from a certain lower limit, a, to an upper limit, b. Denoting such continuously increasing probabilities by v and the corresponding small probabilities of existence by udv, we have as the total probability of obtaining E from any one of the minor complexes with a productive probability between α and β $(\alpha \geqq a, \beta \leqq b)$

$$p = \int_\alpha^\beta uvdv.$$

The probability that when E has happened it originated from one of those minor complexes, or the probability of existence of some one of those complexes is:

$$P = \frac{\int_\alpha^\beta uvdv}{\int_a^b uvdv}.$$

The situation may be still more simplified by the following considerations. In the continuous total complex between the limits a and b we have altogether situated $(b - a)/dv$ individual minor complexes. If we assume all of these complexes to possess the same probability of existence, we must have:

$$udv = \frac{dv}{b - a}.$$

The two formulas then take on the form:

$$p = \frac{1}{b - a}\int_{\alpha}^{\beta} vdv$$

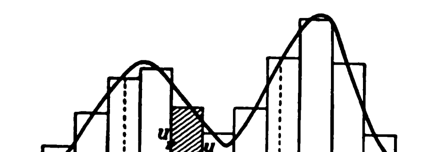
and

$$P = \frac{\int_{\alpha}^{\beta} vdv}{\int_{a}^{b} vdv}.$$

A still more specialized form is obtained by letting $a = 0$ and $b = 1$ which gives:

$$p = \int_{\alpha}^{\beta} vdv \quad \text{and} \quad P = \frac{\int_{\alpha}^{\beta} vdv}{\int_{0}^{1} vdv}.$$

The above formulas may perhaps be made more intelligible to the reader by a geometrical illustration.



Let the various productive probabilities, r, be plotted along the X axis in a Cartesian coördinate system in the interval from a to b $(a < b)$. To any one of these probabilities say v_r there corresponds a certain probability of existence, u_r, represented by a Y ordinate. In the same manner the next following productive probability, v_{r+1}, will have a probability of existence

represented by an ordinate u_{r+1}. It is now possible to represent the various u's by means of areas instead of line ordinates. Thus the probability of existence, u_r, is in the figure represented by the small shaded rectangle, with a base equal to

$$v_{r+1} - v_r = \Delta v_r,$$

and an altitude of u_r, the total area being equal to $\Delta v_r u_r$. That this is so, follows from the well-known elementary theorem from geometry that areas of rectangles with equal bases are directly proportional to their altitudes. The sum of the different u's is thus in the figure represented as the sum areas of the various small rectangles in the staircase shaped histograph. Now according to our assumption v is a continuous function in the interval from a to b. We may, therefore, divide this interval, $b - a$, into n smaller equal intervals. Let

$$v_{r+1} - v_r = \Delta v_r = \frac{b-a}{n}$$

be one of these smaller divisions. By choosing n sufficiently large, $(b - a)/n$ or Δv becomes a very small quantity and by letting n approach infinity as a limiting value we have

$$\lim_{n \doteq \infty} u \frac{b-a}{n} = \lim_{\Delta v \doteq 0} u \Delta v = u dv.$$

In this case the histograph is replaced by a continuous curve and udv is the probability of existence that the productive probability is enclosed between v and $v + dv$.[1]

The probability to get E from any one of the complexes is evidently given by the total area of the small rectangles, or in the continuous case by means of the integral:

$$\int_a^b uvdv\,.$$

[1] A more rigorous analysis would be as follows: We plot along the abscissa axis intervals of the length ϵ so that the middle of the interval has a distance from the origin equal to an integral multiple of ϵ. If now ϵ is chosen sufficiently small, we may regard the probability of existence of u_r for values of the variable v between $r\epsilon - \frac{1}{2}\epsilon$ and $r\epsilon + \frac{1}{2}\epsilon$ as a constant and the probability that v falls between the limits $r\epsilon - \frac{1}{2}\epsilon$ and $r\epsilon + \frac{1}{2}\epsilon$ may hence be expressed as ϵu_r. When ϵ approaches 0 as a limiting value this expression becomes udv. See the similar discussion under frequency curves.

In the same way the probability that E originated from any of the complexes between α and β is:

$$\frac{\int_\alpha^\beta uvdv}{\int_a^b uvdv}.$$

The special case $a = 0$ and $b = 1$ needs no further commentary.

We are now in a position to consider the examples of Bing and Kroman. Any student familiar with multiple integration will find no difficulty in the following analysis. For the benefit of readers to whom the evaluation of the various integrals may seem somewhat difficult, we may refer to the addenda at the close of this treatise or to any standard treatise on the calculus as, for instance, Williamson's "Integral Calculus."

48. Example 24.—An urn contains a very large number of similarly shaped balls. In 10 successive drawings (with replacements) we have obtained 7 with the number 1, 2 with the number 2, and one having the number 3. What is the probability to obtain a ball with another number in the following drawing?

We must here distinguish between 4 kinds of balls, namely balls marked 1, 2, 3, or "other balls." A general scheme of distribution of the balls in the urn may be given through the following scheme:

nx balls marked with the number 1,
ny " " " " " 2,
nz " " " " " 3 and
$nt = n(1 - x - y - z)$ other balls.

Here x, y, z and t represent the respective productive probabilities. If we now let all such probabilities assume all possible values between 0 and 1 with intervals of $1/n$, we obtain the possible conditions in the total complex of actions. Each of these conditions has a probability of existence, s, and the productive probabilities x, y, z, and $1 - x - y - z$. The original probability for 7 ones, 2 twos and 1 three in 10 drawings is:

$$P = \frac{10!}{7!2!1!}\Sigma s \cdot x^7 \cdot y^2 \cdot z.$$

Now when n is a very large number the interval $1/n$ becomes a very small quantity, and we may approximately write:

$$s = udxdydz,$$

and also write the above sum as a triple integral:

$$P = \frac{10!}{7!2!1!}\int_0^1\int_0^p\int_0^q u \cdot x^7 \cdot y^2 \cdot z \cdot dx \cdot dy \cdot dz,$$

where

$$p = 1 - x \quad \text{and} \quad q = 1 - x - y.$$

If now the above event has happened, then the probability to get a different marked ball in the 11th drawing is:

$$Q = \frac{\int_0^1\int_0^p\int_0^q u \cdot x^7 \cdot y^2 \cdot z(1 - x - y - z) \cdot dx \cdot dy \cdot dz}{\int_0^1\int_0^p\int_0^q u \cdot x^7 \cdot y^2 \cdot z \cdot dx \cdot dy \cdot dz}.$$

It is, however, quite impossible to evaluate the above integral without knowing the form of the function u; but unfortunately our information at hand tells us absolutely nothing in regard to this. Perhaps the balls bear the numbers 1, 2 and 3 only, or perhaps there is an equal distribution up to 10,000 or any other number. Our information is really so insufficient that it is quite hopeless to attempt a calculation of the a posteriori probability.

Many adherents of the inverse probability method venture, however, boldly forth with the following solution based upon the perfectly arbitrary hypothesis that all the u's are of equal magnitude. This gives the special integral:

$$Q = \frac{\int_0^1\int_0^p\int_0^q x^7 \cdot y^2 \cdot z(1 - x - y - z), dx \cdot dy \cdot dz}{\int_0^1\int_0^p\int_0^q x^7 \cdot y^2 \cdot z \cdot dx \cdot dy \cdot dz}$$

where once more it must be remembered that

$$x + y + z \leqq 1.$$

In this case the limits of x are 0 and 1, those of y are 0 and $1 - x$ and those of z are 0 and $1 - x - y$.

This is a well-known form of the triple integral which may be evaluated by means of Dirichlet's Theorem:

$$U = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} x^{b-1} y^{m-1} z^{n-1} dx \cdot dy \cdot dz = \frac{\Gamma(b)\Gamma(m)\Gamma(n)}{\Gamma(1+b+m+n)}.$$

(See Williamson's Calculus.)

Remembering the well-known relation between gamma functions and factorials, viz. $\Gamma(n+1) = n!$, we find by a mere substitution in the integral, the value of the probability in question to be 1:14. Another and equally plausible result is obtained by a slightly different wording of the problem.

Ten successive drawings have resulted in balls marked 1, 2, or 3. What is the probability to obtain a ball not bearing such a number in the 11th drawing? This probability is given by the formula.

$$\frac{\int_0^1 v^{10}(1-v)dv}{\int_0^1 v^{10}dv} = 1:12.$$

Quite a different result from the one given above.

49. Example 25—Bing's Paradox.—A still more astonishing paradox is produced by Bing when he gives an example of Bayes's Rule to a problem from mortality statistics. A mortality table gives the ratio of the number of persons living during a certain period, to the number living at the beginning of this period, all persons being of the same age. By recording the deaths during the specified period (say one year) it has been ascertained that of s persons, say forty years of age at the beginning of the period, m have died during the period. The observed ratio is then $(s-m)/s$. If s is a very large number this ratio may (as we shall have occasion to prove at a later stage) be taken as an approximation of the true ratio of probability of survival during the period. If s is not sufficiently large the believers in the inverse theory ought to be able to evaluate this ratio by an application of Bayes's Rule, by means of an analysis similar to the one as follows:

Let y be the general symbol for the probability of a forty-year-old person being alive one year from hence. Each of such persons will in general be subject to different conditions, and the general symbol, y, will therefore have to be understood as the

symbol for all the possible productive probability values changing from 0 to 1 by a continuous process.

Assuming s a very large number each condition will have a probability of existence equal to udy. We may now ask: What is the probability that the rate of survival of a group of s persons aged 40 is situated between the limits α and β?

The answer according to Bayes's Rule is:

$$\frac{\int_{\alpha}^{\beta} y^{s-m}(1-y)^m u\,dy}{\int_0^1 y^{s-m}(1-y)^m u\,dy}. \tag{I}$$

Let us furthermore divide the whole year into two equal parts and let y_1 be the probability of surviving the first half year, y_2 the probability of surviving the second half, and $u_1 \cdot dy_1$, $u_2 \cdot dy_2$ the corresponding probabilities of existence. Then the respective a posteriori probabilities for y_1 and y_2 are:

$$\frac{y_1^{s-m_1}(1-y_1)^{m_1}\,u_1dy_1}{\int_0^1 y_1^{s-m_1}(1-y_1)^{m_1}u_1dy_1}$$

and

$$\frac{y_2^{s-m}(1-y_2)^{m_2}u_2dy_2}{\int_0^1 y_2^{s-m}\,(1-y_2)^{m_2}u_2dy_2} \qquad (m_1 + m_2 = m)$$

(m_1 and m_2 represent the number of deaths in the respective half years.) The probability that both y_1 and y_2 are true is then according to the multiplication theorem:

$$\frac{y_1^{s-m_1}(1-y_1)^{m_1}u_1dy_1y_2^{s-m}(1-y_2)^{m_2}u_2dy_2}{\int_0^1 y_1^{s-m_1}(1-y_1)^{m_1}u_1dy_1\int_0^1 y_2^{s-m}(1-y_2)^{m_2}u_2dy_2}$$

where $y = y_1 \cdot y_2$.

The probability that the probability of survival for a full year, y, is situated between the limits α and β is therefore:

$$\frac{\int\int y_1^{s-m_1}(1-y_1)^{m_1}y_2^{s-m}(1-y_2)^{m_2}u_1 \cdot u_2 \cdot dy_1 \cdot dy_2}{\int_0^1 y_1^{s-m_1}(1-y_1)^{m_1}u_1dy_1\int_0^1 y_2^{s-m}(1-y_2)^{m_2}u_2dy} \tag{II}$$

where the limits in the double integral in the numerator are determined by the relation:

$$\alpha \leqq y_1 y_2 \leqq \beta.$$

Choosing the principle of insufficient reason as the basis of our calculations, merely assuming that all possible events are, in the absence of any grounds for inference, equally likely, the various quantities expressed by the general symbol, u, become equal and constant and cancel each other in numerator and denominator, which brings the a posteriori probabilities expressed by (I) and (II) to the forms:

$$\frac{\int_\alpha^\beta y^{s-m}(1-y)^m dy}{\int_0^1 y^{s-m}(1-y)^m dy} \tag{III}$$

and

$$\frac{\int\int y_1^{s-m_1}(1-y_1)^{m_1} y_2^{s-m_1-m_2}(1-y_2)^{m_2} dy_1 \cdot dy_2}{\int_0^1 y_1^{s-m_1}(1-y_1)^{m_1} \int_0^1 y_2^{s-m}(1-y_2)^{m_2} dy_1 \cdot dy_2} \tag{IV}$$

where the limits in the numerator in the latter expression are determined by the relation : $\alpha < y_1 y_2 < \beta$.

Letting

$$y_2 = \frac{y}{y_1}$$

and then

$$1 - y_1 = z(1-y)$$

this latter expression may after a simple substitution be brought to the form:

$$\frac{\int_\alpha^\beta y^{s-m}(1-y)^{m+1} dy \int_0^1 \frac{z^{m_1}(1-z)^{m_2} dz}{1 - z(1-y)}}{\int y_1^{s-m_1}(1-y_1)^{m_1} dy_1 \int_0^1 y_2^{s-m}(1-y_2)^{m_2} dy_2}. \tag{V}$$

(See appendix.)

Mr. Bing now puts the further question: What is the probability that a new person forty years of age, entering the original large

group of s persons, will survive one year, when we assume $m_1 = m_2 = 0$? (III) gives the answer:

$$\frac{\int_0^1 y^{s+1}dy}{\int_0^1 y^s dy} = \frac{s+1}{s+2}.$$

Formula (V), on the other hand, gives us:

$$\frac{\int_0^1 y^{s+1}(1-y)dy \int_0^1 \frac{dz}{1-z(1-y)}}{\int_0^1 y_1^s dy_1 \int_0^1 y_2^s dy_2} = \left(\frac{s+1}{s+2}\right)^2.$$

As the above analysis is perfectly general, we might equally well have applied it to each of the semi-annual periods, which would give us an a posteriori probability of survival equal to $\left(\frac{s+1}{s+2}\right)^2$ for each half year, or a compound probability of $\left(\frac{s+1}{s+2}\right)^4$ for the whole year. Extending this process it is easily seen that by dividing the year into parts, we shall have $\left(\frac{s+1}{s+2}\right)^n$ as the final probability a posteriori that a forty-year-old person will reach the age of forty-one. By letting n increase indefinitely the above quantity approaches 0 as its limiting value and we obtain thus the paradox of Bing:

If, among a large group of s equally old persons, we have observed no deaths during a full calendar year then another person of the same age outside the group is sure to die inside the calendar year.

This is evidently a very strange result, and yet, working on the basis of the principle of insufficient reason, the mathematical deductions and formula exhibit no errors.

Mr. Bing disposes of the whole matter by simply denying the validity and existence of a posteriori probabilities. Dr. Kroman on the other hand defends Bayes's Rule. "Mathematics," Kroman says, "is—as Huxley has justly remarked—an exceedingly fine mill stone, but one must not expect to get wheat flour after having put oats in the quern." According to the

Danish scholar the paradox is due to the use of a wrong formula. We ought to have used the general formula (II) instead of formula (V) which is a special case. In the general formula we encounter the functions u, denoting the probability existence of the various productive probabilities y. As we do not know anything about this function u it is hopeless to attempt a calculation. This brings the criticism down to the fundamental question whether we shall build the theory of probabilities on the principle of "cogent reason" or the principle of "insufficient reason."

50. Conclusion.—Contradictory results of a similar kind to the ones given above have led several eminent mathematicians to a complete denunciation of the laws underlying a posteriori probabilities. Professor Chrystal, especially, becomes extremely severe in his criticism in the previously mentioned address before the Actuarial Society of Edinburgh. He advises "practical people like the actuaries, much though they may justly respect Laplace, not to air his weaknesses in their annual examinations. The indiscretions of a great man should be quietly allowed to be forgotten." Although one may heartily agree with Professor Chrystal's candid attack on the belief in authority, too often prevailing among mathematical students, I think—aside from the fact that the rule was originally given by Bayes—that the great French savant has been accused unjustly as the following remarks perhaps may tend to show.

In our statement of Bayes's Rule, we followed an exact mathematical method, and the final formula (I) is theoretically as correct as any previously demonstrated in this work. The customary definition of a mathematical probability as the ratio of equally favorable to coördinated possible cases, is not done away with in this new kind of probabilities; the former are found in the numerator and the latter in the denominator; and if we take care that each of the particular formulas, with its definite requirements, is applied to its particular case, we do not go beyond pure mathematics or logic. But are we able to get complete and exact information about these requirements? In the example of the tossing of a coin with two heads, this information was at hand. Here we were able to enumerate exactly the different mutually exclusive causes from which the observed

event originated. We were also able to determine the exact quantitative measures for the probabilities, κ, that these complexes existed as well as the different productive probabilities, ω. Here the most rigid requirements could be satisfied, and the rule gave therefore a true answer.

In the other examples we encountered a different state of affairs. Here we were not able to enumerate directly the different complexes of causes from which the event originated, but were forced to form different and arbitrary hypotheses about the complexes of origin, F, and each hypothesis gave, in general, a different result. Furthermore, we assumed a priori that the different probabilities of the actual existence of the complexes were all equal in magnitude, and it was, therefore, the special formula (II) we employed in the determination of the a posteriori probabilities. In this formula, the different κ's do not enter at all as a determining factor; only the productive probabilities, ω, are considered. The assumption that all the κ's are equal in magnitude is based upon the principle of insufficient reason, or as Boole calls it, " the equal distribution of ignorance."

The principle of equal distribution of ignorance makes in the case of continuously varying productive probabilities, v, the function, u, of the probabilities of existence of the various complexes equal to a constant quantity. In other words, the curve in Fig. 1, is replaced by a straight line of the form, $u = k$. Now, as a matter of fact, we possess in most cases, some partial knowledge of the complexes of action producing the event in question. This partial knowledge—although far from complete enough to make a rigorous use of formula (I)—is nevertheless sufficient to justify us in discarding completely any general hypothesis assuming such simple conditions as above. Such partial knowledge is, for instance, found in the Paradox of Bing. Here the rather absurd hypothesis was made that the possible values of the probability of surviving a certain period were equally probable. In other words, it is equally probable that there will die 0, 1, 2, $\cdots$, or s persons in the particular period. " Common sense, however, tells us that it is far more probable that, for instance, 90 per cent. of a large number of forty-year-old persons will survive the period than no one or every one will die

in the same period " (Kroman). The indiscreet use of formula (II) therefore naturally leads to paradoxical results. On the other hand, the fallacy of the happy-go-lucky computers, employing the special case (II) of Bayes's Rule, as well as the critics of Laplace, lies in their failure to make a proper distinction between " equal distribution of ignorance " and " partial cogent reason," which latter expression properly may be termed " an unequal distribution of ignorance." If, despite the actual presence of such unequal distribution of ignorance, we still insist in using the special formula (II), which is only to be used in the case of an equal distribution of ignorance, it is no wonder we encounter ambiguous answers. Not the rule itself, its discoverer, or Laplace, but the indiscreet computer is the one to blame. Messrs. Bing, Venn and Chrystal, in their various criticisms, have filled the quern with some rather "wild oats" and expected to get wheat flour; and that one of those critics in his disappointment in not getting the expected flour should blame Laplace, is hardly just.

So much for the principle of " equal distribution of ignorance." It may be of interest to see how matters turn out when we like von Kries insist upon the principle of " cogent reason " as the true basis of our computations. The reader will quite readily see that a rigorous application of the Rule of Bayes in its most general form as given by formula (I) really tacitly assumes this very principle. In formula (I), we require not alone an exact enumeration of the various complexes from which the observed event may originate, but also an exact and complete information about the structure of such complexes in order to evaluate their various probabilities of existence. If such information is present, we can meet even the most stringent requirements of the general formula, and we will get a correct answer. But in the vast majority of cases, not to say all cases, such information is not at hand, and any attempt to make a computation by means of Bayes's Rule must be regarded as hopeless. We may, however, again remark that very seldom we are in complete ignorance of the conditions of the complexes, which is the same thing as saying that we are not in a position to employ the principle of equal distribution of ignorance in a rigorous manner. From

other experiments on the same kind of event, or from other sources, we may have attained some partial information, even if insufficient to employ the principle of cogent reason. Is such information now to be completely ignored in an attempt to give a reasonable, although approximate answer? It is but natural that the mathematician should attempt to obtain as much of such information as possible and use it in the evaluation of the various probabilities of existence. Thus for instance, if, in the Paradox of Bing, we had observed that the probability of survival for a forty-year-old person never had been below .75 and never above .95, it would be but reasonable to substitute those limits in their proper integrals in order to attain an approximate answer. To illustrate this somewhat subjective determination of an a posteriori probability, we take another example from the memoirs of Bing and Kroman.

Example (24).—A merchant receives a cargo of 100,000 pieces of fruit. If every single fruit is untainted, the value of the cargo may be put at 10,000 Kroner. On the other hand, any part of the cargo more or less tainted is considered worthless. The merchant has never before received a similar cargo and does not know how the fruit has been affected by travel. As samples, he has selected 30 pieces picked at random from the cargo and all samples proved to be fresh. He asks a mathematician what value he can put on the cargo.

If the mathematician uses the special formula (II), assuming an equal distribution of ignorance, therefore assuming that it is equally probable that for example none, 5,000 or all the individual pieces of fruit were untainted, the answer is:

$$10{,}000\frac{\int_0^1 v^{31}dv}{\int_0^1 v^{30}dv} = 9687.5 \text{ Kroner.}$$

If we use the true rule, the a posteriori probability of the wholesomeness of the cargo is given by the integral:

$$\frac{\int_0^1 uv^{31}dv}{\int_0^1 uv^{30}dv}$$

where v is the general expression for a possible probability of wholesomeness between 0 and 1 and udv the corresponding probability of existence. Now if the mathematician has no complete information as to this particular function, u, it would be foolish of him to attempt a calculation, since the hypothesis of an equal probability of existence for all possible values of v evidently gives an arbitrary and perhaps a very erroneous result. On the other hand, the computer may possibly have access to some partial information. Perhaps the merchant has received fruit of a similar kind or heard about cargoes of this particular kind of fruit received by other dealers. If now the merchant were able to inform the computer that in a great number of similar cases the probability of wholesomeness had been between 0.9 and 1 with an approximately even distribution, while it never had been below 0.9, then nothing would hinder the mathematician to present the following computation:

$$\frac{\int_{0.9}^{1} v^{31}dv}{\int_{0.9}^{1} v^{30}dv} = 0.9726$$

and tell the merchant that on the basis of the information given 9,726 Kroner would be a fair price for the cargo.

This is really the point of view taken by the English mathematician, Professor Karl Pearson, one of the ablest writers on mathematical statistics of the present time, when he says: "I start, as most writers on mathematics have done, with 'the equal distribution of ignorance' or I assume the truth of Bayes's Theorem. I hold this theorem not as rigidly demonstrated, but I think with Edgeworth that the hypothesis of the equal distribution of ignorance is, within the limits of practical life, justified by our experience of statistical ratios, which are unknown, i. e., such ratios do not tend to cluster markedly round any particular point."

To sum up the above remarks: Theoretically Bayes's Rule is true. If we are able to enumerate and determine the probabilities of existence of the complexes of origin it will also give true results in practice. If we are justified in assuming the principle

of "insufficient reason" or "equal distribution of ignorance" as the basis for our calculations, formula (II) may be employed with exact results after a rigid enumeration of the complexes. If the principle of "cogent reason" is required as the basis, an exact computation is in general hopeless, and we can only after having obtained partial subjective information give an approximate answer.

With these remarks we shall conclude the elementary discussion of the merely theoretical part of the subject. The following chapters require in most cases a knowledge of the infinitesimal calculus, and many of the questions discussed above will appear in a new and instructive light by this treatment.

CHAPTER VII.

THE LAW OF LARGE NUMBERS.

51. A Priori and Empirical Probabilities.—In the previous chapters we limited ourselves to the discussion of such mathematical probabilities, where we, a priori, on account of our knowledge of the various domains or complexes of actions, were able to enumerate the respective favorable and unfavorable possibilities associated with the occurrence or non-occurrence of the event in question. "The real importance of the theory of probability in regard to mass phenomena consists, however, in determining the mathematical relations of the various probabilities not in a deductive, but in an empirical manner—without an a priori exhaustive knowledge of the mutual relations and actions between cause and effect—by means of statistical enumeration of the frequency of the observed event. The conception of a probability finds its justification in the close relation between the *mathematical probabilities* and *relative frequencies* as determined in a purely empirical way. This relation is established by means of the famous Law of Large Numbers" (A. A. Tschuprow).

To return to our original definition of a mathematical probability as the ratio of the favorable to the coördinated equally possible cases, we first notice that this definition is wholly arbitrary like many mathematical definitions. The contention of Stuart Mill that every definition contains an axiom is rather far stretched. In mathematics a definition does not necessarily need to be metaphysical. A striking example is offered in mechanics by the definitions of force as given by Lagrange and Kirchhoff. What is force? "Force," Lagrange says, "is a cause which tends to produce motion." Kirchhoff on the other hand tells us that force is the product of mass and acceleration. Lagrange's definition is wholly metaphysical. Whenever a definition is to be of use in a purely exact science such as mathematics, it must teach us how to measure the particular phenomena which we are investigating. Thus, to quote Poincaré,

"it is not necessary that the definition tells us what force really is, whether it is a cause or the effect of motion."

An analogous case is offered in the criticism of a mathematical probability as defined by Laplace, and the attempts to place the whole theory of probabilities on a purely empirical basis by Stuart Mill, Venn and Chrystal. These writers contend "that probability is not an attribute of any particular event happening on any particular occasion. Unless an event can happen, or be conceived to happen a great many times, there is no sense in speaking of its probability." The whole attack is directed against the definition of a mathematical probability in a *single trial* which definition, evidently by the empiricists, is regarded as having no sense. The word "sense" must evidently be considered as having a purely metaphysical meaning. In the same manner Kirchhoff's definition might be dismissed as having no sense, since it would seem as difficult to conceive force as a purely mathematical product of two factors, mass and acceleration, as it is to conceive the definition of a mathematical probability as a ratio.

The metaphysical trend of thought of the above writers is shown in their various definitions of the probability of an event. Mill defines it merely as the relative frequency of happenings inside a large number of trials, and Venn gives a similar definition, while Chrystal gives the following:

"If, on taking any very large number N out of a series of cases in which an event, E, is in question, E happens on pN occasions, the probability of the event, E, is said to be p."

Let us, for a moment, look more closely into these statements. Any definition, if it bears its name rightly, must mean the same to all persons. Now, as a matter of fact, the vagueness in a half metaphorical term like "any very large number" illustrates its weakness. The question immediately confronts us "what is a very large number?" Is it 100, 1,000 or perhaps 1,000,000?

A fixed universal standard for the value of N seems out of the question and the definition—although perhaps readily grasped in a "general way"—can hardly be said to be happily chosen.

Another, and perfectly rigorous definition, is the following one given by the Danish astronomer and actuary, T. N. Thiel·

Thiele tells us that "common usage" has assigned the word probability as the name "for the limiting value of the relative frequency of an event, when the number of observations (trials), under which the event happens, approach infinity as a limit." A similar definition is later on given by the American actuary R. Henderson, who says: "The numerical measure which has been universally adopted for the probability of an event under given circumstances is the ultimate value, as the number of cases is indefinitely increased, of the ratio of the number of times the event happens under those circumstances to the total possible number of times." There is nothing ambiguous or vague in these definitions. Infinity, taken in a purely quantitative sense, has a perfectly uniform meaning in mathematics. The new definition differs, however, radically from our customary definition of a mathematical a priori probability. We cannot, therefore, agree with Mr. Henderson when he continues "the measure there given has been universally adopted and this holds true in spite of the fact that the rule has been stated in ways which on their face differ widely from that above given. The one most commonly given is that if an event can happen in *a* ways and fail in *b* ways all of which are equally likely, the probability of the event is the ratio of *a* to the sum of *a* and *b*. It is readily seen that if we read into this statement the meaning of the words "equally likely," this measure, so far as it goes, reduces to a particular case of that given above."

In order to investigate this statement somewhat more closely, let us try to measure the probability of throwing head with an ordinary coin by both our old definition of a mathematical probability and the definition by Mr. Henderson of what we shall term an empirical probability. Denoting the first kind of probability by $P(E)$ and the second by $P'(E)$ we have in ordinary symbols

$$P(E) = \tfrac{1}{2}$$

$$P'(E) = \lim_{v=\infty} F(E, v)$$

where the symbol $F(E, v)$ denotes the relative frequency of the event, E, in v total trials. No a priori knowledge will tell us offhand if $P'(E)$ will approach $\tfrac{1}{2}$ as its ultimate value. The

two methods are radically different. By the first method the determination of the numerical measure of a probability depends simply on our ability to judge and segregate the equally possible cases into cases favorable and unfavorable to the event E. By the second method the determination of the probability depends, not alone on the segregation and consequent enumeration of the favorable from the total cases, but chiefly on the extent of our observations or trials on the event in question.

52. Extent and Usage of Both Methods.—Before entering into a more detailed discussion of the actual quantitative comparison of the two methods, it might be of use to compare their various extent of usage. In this respect the empirical method is vastly superior to the a priori. A rigorous application of the a priori method, as far as concrete problems go, is limited to simple games of chance. As soon as we begin to tackle sociological or economical practical problems it leaves us in a helpless state. If we were to ask about the probability that a certain person forty years of age would die inside a year, it would be of little use to try to determine this in an a priori manner. Even a purely deductive process, as illustrated by Bayes's Rule in the earlier chapters, leads to paradoxical results. Our a priori knowledge of the complexes of causes governing death or survival is so incomplete that even a qualitative—not to speak of a quantitative—judgment is out of the question. The empirical method shows us at least a way to obtain a measure for the probability of the event in question. By observing during a period of a year an infinite number of forty-year-old persons of whom, after an exhaustive qualitative investigation, we are led to believe that their present conditions as far as health, social occupation, environments, etc., are concerned are equally similar, we may by an enumeration of those who died during the year obtain the desired ratio as defined by $P'(E)$. Of course, observation an infinite number is practically impossible. An approximate ratio may be formed by taking a finite, but a large, number of cases under observation. But how large a number? This very question leads straightforward to another problem, namely the quantitative determination of the range of variance between the approximate ratio and the ideal ultimate ratio as defined

the relation

$$P'(E) = \lim_{v=\infty} F(E, v).$$

Since it is impossible to make an infinite number of observations we cannot find the exact value of the range of such variations. But we may, however, determine the probability that this range does not exceed a certain fixed quantity, say λ, in absolute magnitude. Stated in compact form our problem reduces to the following form: To determine the probability of the existence of the following inequality:

$$\left| \lim_{v=\infty} F(E, v) - \frac{\alpha}{s} \right| \leqq \lambda$$

where both α and s are finite numbers. This, to a certain extent, contains in a nut shell some of the most important problems in probabilities.

The above problem may be solved in two distinct ways. The first, and perhaps the most logical way, is by a direct process. This is the method followed by T. N. Thiele in his " Almindelig Iagttagelseslære,"[1] published in Copenhagen, 1889, a most original work, which moves along wholly novel lines. Thiele distinguishes between (1) Actual observation series as recorded from observation, in other words statistical data. (2) Theoretical observation series giving the conclusions as to the outcome of future observations and (3) Methodical laws of series where the number of observations is increased indefinitely. By such a process, purely a theory of observations, the whole theory of probability becomes of secondary importance and rests wholly upon the theory of observed series, a fact thoroughly emphasized by Thiele himself. When the author first, in the closing chapters of his book, makes use of the word probability it is only because " common usage " has assigned this word as the name for the ultimate frequency ratio designated by our symbol $\lim_{v=\infty} F(E, v)$.

The problem may, however, be solved in an indirect way, which is the one I shall adopt. This method, as first consistently deduced by Laplace, has for its basis our original definition of a mathematical a priori probability and may be briefly sketched as follows: We first of all postulate the existence of an a priori

[1] English edition, "Theory of Observations," London, 1905.

probability as defined, although its actual determination, by a priori knowledge, is impossible except in a few cases, as, for instance, simple games of chance, drawing balls from urns, etc. Denoting such a probability by $P(E)$, or p, we next ask, What will be the expected number, say α, of actual happenings of the event, E, expressed in terms of s and p, when we make s consecutive trials instead of a single trial, and what will be the number of happenings of E when s approaches infinity as its ultimate value? If such a relation is found between p, α and s, where p is the unknown quantity, we have also found a means of determining the value of p in known quantities. Our next question is— What is the probability that the absolute value of the difference between p and the relative frequency of the event as expressed by the ratio of α to s does not exceed a previously assigned quantity? Or the probability that

$$\left| p - \frac{\alpha}{s} \right| \leqq \lambda?$$

Now, as the reader will see later, we shall prove that

$$\lim_{v=\infty} F(E, v) = P(E) = p.$$

It must, however, be remembered that this result is reached by a mathematical deduction, based upon the postulate of mathematical probabilities, and not in the manner as suggested in the above statement by Mr. Henderson.

It is only after having established such purely quantitative relations that we are entitled to extend the laws of mathematical probabilities as deduced in the earlier chapters to other problems than the simple problems of games of chance.

53. Average a Priori Probabilities.—In the previous paragraphs of this chapter, another important matter is to be noted, namely the assumption that the complex of causes producing the event in question remains constant during the repeated trials (observations), or, stated in other words the mathematical a priori probability remains constant. Under this limitation the extension of the laws of mathematical probabilities would have but a very limited practical application. In all statistical mass phenomena such an ideal state of affairs is rather a very

rare exception. If we consider an ordinary mortality investigation we know with absolute certainty that no two persons are identically alike as far as health, occupation, environment and numerous other things are concerned. Thus the postulated mathematical probability for death or survival during a whole calendar year will in general be different for each person. We may, however, conceive an average probability of survival for a full year defined by the relation

$$p_0 = \frac{p_1 + p_2 + p_3 + \cdots p_s}{s} = \frac{\Sigma p}{s},$$

where $p_1, p_2, p_3, \cdots$ are the postulated probabilities of each individual under observation. Our task is now to find:

1. An algebraic relation between the average probability as defined above, the absolute frequency α and the total number of observations (trials) s,
2. The same relation when s approaches α as its ultimate value,
3. The probability of the existence of the inequality,

$$\left| p_0 - \frac{\alpha}{s} \right| \leqq \lambda,$$

where α denotes the absolute frequency of the occurrence of the event, s the total number of observations (trials) and λ an arbitrary constant.

54. The Theory of Dispersion.—As we mentioned before the empirical ratio α/s represents only an approximation of the ideal ultimate value of $\lim_{v=\infty} F(E, v)$. If we now make a series of observations (trials) on the occurrence of a certain event E, such that instead of a single set of observations of s individual observations we take N such sets, we shall have N relative frequency ratios:

$$\frac{\alpha_1}{s}, \frac{\alpha_2}{s}, \frac{\alpha_3}{s}, \cdots \frac{\alpha_N}{s}.$$

Since the ratios are approximations only of the ultimate ratio they will in general exhibit discrepancies as to their numerical values and may be regarded as N different empirical approximations. The question now arises how these various empirical ratios group themselves around the value of $\lim_{v=\infty} F(E, v)$. The dis-

tribution of the empirical ratios around the ultimate ratio is by Lexis called "dispersion."

55. Historical Development of the Law of Large Numbers.— The first mathematician to investigate the problems we have roughly outlined in the previous paragraphs was the renowned Jacob Bernoulli in the classic, "Ars Conjectandi," which rightly may be classified as one of the most important contributions on the subject. Bernoulli's researches culminate in the theorem which bears his name and forms the corner-stone of modern mathematical statistics. That Bernoulli fully realized the great practical importance of these investigations is proven by the heading of the fourth part of his book which runs as follows: "Artis Conjectandi Pars Quarta, tradens usum et applicationem præcedentis doctrinæ in civilibus et œconomicis." It is also here that we first encounter the terms "a priori" and "a posteriori" probabilities. Bernoulli's researches were limited to such cases where the a priori probabilities remained constant during the series or the whole sets of series of observations. Poisson, a French mathematician, treated later in a series of memoirs the more general case where the a priori probabilities varied with each individual trial. He also introduced the technical term, "Law of Large Numbers" ("Loi des Grand Nombres"). Finally Lexis through the publication in 1877 of his brochure, "Zur Theorie der Massenerscheinungen der menschlichen Gesellschaft," treated the dispersion theory and forged the closing link of the chain connecting the theory of a priori probabilities and empirical frequency ratios. Of late years the Russian mathematician, Tchebycheff, the Scandinavian statisticians, Westergaard and Charlier, and the Italian scholar, Pizetti, have contributed several important papers. It is on the basis of these papers that the following mathematical treatment is founded. In certain cases, however, we shall not attempt to enter too deeply into the theory of certain definite integrals, which is essential for a rigorous mathematical analysis, but which also requires an extensive mathematical knowledge which many of my readers, perhaps, do not possess. To readers interested in the analysis of the various integrals we may refer to the original works of Czuber and Charlier.

CHAPTER VIII.

INTRODUCTORY FORMULAS FROM THE INFINITESIMAL CALCULUS.

56. Special Integrals.—In the following chapters we shall attempt to investigate the theory of probabilities from the standpoint of the calculus. Although a knowledge of the elements of this branch of mathematics is presupposed to be possessed by the student, we shall for the sake of convenience briefly review and demonstrate a few formulas from the higher analysis of which we shall make frequent use in the following paragraphs. All such formulas have been given in the elementary instruction of the calculus, and only such readers who do not have this particular branch of mathematics fresh in memory from their school days need pay any serious attention to the first few paragraphs.

57. Wallis's Expression for π as an Infinite Product.—We wish first of all to determine the value of the definite integral:

$$J_n = \int_0^{\pi/2} \sin^n x dx, \tag{1}$$

under the assumption that n is a positive integral number. This integral is geometrically equal to the area between the x axis, the axis of y, the ordinate corresponding to the abscissa $\frac{1}{2}\pi$ and the graph of the function $y = \sin^n x$. Letting $u' = D_x u = \sin x$, $v = \sin^{n-1} x$, we get by partial integration:

$$J_n = - \cos x \sin^{n-1} x \Big]_0^{\pi/2} + \int_0^{\pi/2} \cos x(n-1) \sin^{n-2} x \cos x dx. \tag{2}$$

If we substitute the upper and lower limits in the first term on the right hand side of the above expression for J_n this term reduces to 0, assuming $n > 1$. Thus we have:

$$J_n = (n - 1)\int_0^{\pi/2} \sin^{n-2} x \cdot \cos^2 x dx.$$

Putting $\cos^2 x = 1 - \sin^2 x$, we get:

$$J_n = (n-1)\int_0^{\pi/2} \sin^{n-2} x dx - (n-1)\int_0^{\pi/2} \sin^n x dx. \tag{3}$$

The last integral is, however, equal to J_n and the first integral is, following the notation from (1), equal to J_{n-2}. We shall therefore have:

$$J_n + (n-1)J_n = (n-1)J_{n-2},$$

or

$$nJ_n = (n-1)J_{n-2}. \tag{4}$$

Replacing n by $n-1$, $n-2$, $n-3$, $\cdots$ successively we get:

$$\begin{aligned} nJ_n &= (n-1)J_{n-2}, \\ (n-1)J_{n-1} &= (n-2)J_{n-3}, \\ (n-2)J_{n-2} &= (n-3)J_{n-4}, \\ &\cdots\cdots \\ &\cdots\cdots \end{aligned}$$

According as n is even or uneven we shall have one of the following equations at the bottom of the recursion formula:

$$J_0 = \int_0^{\pi/2} \sin^0 x dx = \int_0^{\pi/2} dx = \tfrac{1}{2}\pi,$$

or

$$J_1 = \int_0^{\pi/2} \sin x dx = -\cos x\Big]_0^{\pi/2} = 1. \tag{5}$$

If, for even values of n, we let $n = 2m$, and, for uneven values, $n = 2m-1$, we get finally the following recursion formulas:

$$\begin{aligned} 2mJ_{2m} &= (2m-1)J_{2m-2}, & (2m-1)J_{2m-1} &= (2m-2)J_{2m-3}, \\ (2m-2)J_{2m-2} &= (2m-3)J_{2m-4}, & (2m-3)J_{2m-3} &= (2m-4)J_{2m-5}, \\ &\cdots\cdots & &\cdots\cdots \\ &\cdots\cdots & &\cdots\cdots \\ 2J_2 &= 1\cdot\tfrac{1}{2}\pi, & 3J_3 &= 2\times 1. \end{aligned}$$

Successive multiplication of the above equations gives us finally:

$$\begin{aligned} J_{2m} &= \frac{(2m-1)(2m-3)\cdots 1}{2m(2m-2)\cdots 2} \times \frac{\pi}{2}, \\ J_{2m-1} &= \frac{(2m-2)(2m-4)\cdots 2}{(2m-1)(2m-3)\cdots 3}. \end{aligned} \tag{6}$$

We may now draw some very interesting conclusions from the

above equations. Both integrals represent geometrically areas bounded by the graphs of the functions:

$$y = \sin^{2m} x \text{ and } y = \sin^{2m-1} x \text{ respectively.}$$

The difference of the ordinates of these graphs, namely:

$$(\sin x - 1) \sin^{2m-1} x$$

is evidently decreasing with increasing values of the positive integer n, since $\sin x$ lies between 0 and $+ 1$ and $\sin^{2m-1} x$ approaches the value 0 except for certain values of x. The larger we select m the less is the difference of the two areas and the ratio will therefore approach 1, or the expression

$$\frac{(2m-2)(2m-4)\cdots 2}{(2m-1)(2m-3)\cdots 3} \div \frac{(2m-1)(2m-3)\cdots 3}{2m(2m-2)\cdots 2} = \frac{\pi}{2}.$$

Hence:

$$\frac{\pi}{2} = \lim_{m=\infty} \frac{2^2 \cdot 4^2 \cdot 6^2 \cdots (2m-2)^2 \cdot 2m}{1^2 \cdot 3^2 \cdot 5^2 \cdots (2m-3)^2 (2m-1)^2}.$$

Multiplying with $2^2 \cdot 4^2 \cdot 6^2 \cdots (2m - 2)^2$ we get:

$$\frac{\pi}{2} = \lim_{m=\infty} \frac{2^{4m-3} m[(m-1)!]^4}{[(2m-1)!]^2} \quad \text{or} \quad \lim_{m=\infty} \frac{2^{2m}(m!)^2}{(2m!)\sqrt{2m}} = \sqrt{\pi/2}.$$

This is the formula originally discovered by the English mathematician, John Wallis (1616–1703), and by means of which π may be expressed as an infinite product.

58. De Moivre—Stirling's Formula.—We are now in a position to give a demonstration of Stirling's formula for the approximate value of $n!$ for large values of n. A. de Moivre seems to have been the first to attempt this approximation. In the first edition of his "Doctrine of Chances" (1718) he reaches a result, which must be regarded as final, except for the determination of an unknown constant factor. Stirling succeeded in completing this last step in his remarkable "Methodus Differentialis" (1738). In the second edition of "Doctrine of Chances" (1738) de Moivre gives the complete formula with full credit to Stirling. He mentions as his belief that Stirling in his final calculation possibly has made use of the formula of Wallis. The demonstration by the older English authors is rather lengthy and much shorter

methods have been devised by later writers. Most authors make use of the Eulerian integral of the second order by which any factorial may be expressed by a gamma function:

$$\Gamma(n+1) = \int_0^\infty x^n e^{-x} dx = n!.$$

Another method makes use of the well-known Euler's Summation Formula from the calculus of finite differences. This method is of special interest to actuarial students, who frequently use the Eulerian formula in the computation of various life contingencies. For the benefit of those interested in this particular method we may refer to the treatises of Seliwanoff and Markhoff, two Russian mathematicians.[1]

The Italian mathematician, Cesaro, has, however, derived the formula in a much simpler manner.[2]

Cesaro starts with the inequalities:

$$e < \left(1 + \frac{1}{n}\right)^{n+1/2} < e^{1+\frac{1}{12n(n+1)}}.$$

From a well-known theorem from logarithms we have:

$$\tfrac{1}{2}\log_e \frac{n+1}{n} = \frac{1}{2n+1} + \frac{1}{3(2n+1)^3} + \frac{1}{5(2n+1)^5} + \cdots,$$

which also may be written as follows:

$$N = (n + \tfrac{1}{2})\log_e\left(1 + \frac{1}{n}\right) = 1 + \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^4} + \cdots.$$

If all the coefficients 3, 5, $\cdots$ are replaced by the number 3, we obtain a geometrical series. The summation of this infinite series shows that

$$1 < N < 1 + \frac{1}{12n(n+1)},$$

or

$$e < \left(1 + \frac{1}{n}\right)^{n+1/2} < e^{1+\frac{1}{12n(n+1)}}. \qquad \text{(I)}$$

If we let

$$u_n = \frac{n!\,e^n}{n^{n+1/2}}, \quad u_{n+1} = \frac{(n+1)!\,e^{n+1}}{(n+1)^{n+3/2}},$$

[1] Seliwanoff, "Lehrbuch der Differenzenrechnung," Leipzig, 1905, pages 59–60; Markhoff, "Differenzenrechnung," Leipzig, 1898.

[2] Cesaro, "Corso di analisa algebrica," Torino, 1884, pages 270 and 480.

then

$$\frac{u_n}{u_{n+1}} = \frac{(1 + 1/n)^{n+1/2}}{e}.$$

Dividing the quantities in (I) by e we have:

$$1 < \frac{u_n}{u_{n+1}} < e^{\frac{1}{12n(n+1)}}. \qquad \text{(II)}$$

The exponent of e may be written as follows:

$$\frac{1}{12n(n + 1)} = \frac{1}{12n} - \frac{1}{12(n + 1)}.$$

Making use of this relation (II) may be written in the following form:

$$u_n \cdot e^{-\frac{1}{12n}} < u_{n+1} \cdot e^{-\frac{1}{12(n+1)}} < u_{n+1} < u_n.$$

Denoting the quantity: $u_n \cdot e^{-1/12n}$ by u_n', we shall have two monotone number sequences:

$$u_1,\ u_2,\ u_3, \cdots u_n,\ u_{n+1}, \cdots,$$
$$u_1',\ u_2',\ u_3', \cdots u_n',\ u_{n+1}', \ \cdots.$$

These two sequences show some very remarkable features. With increasing values of n the values of u_n decrease, or the sequence is a monotone decreasing number sequence. The values of u_n' become larger when n is increased and form therefore a monotone increasing number sequence. But any member of this latter series satisfies, however, the inequality

$$u_n' < u_n.$$

Since both number sequences are situated in a finite interval it follows from the well-known theorem of Weierstrass that they both have a clustering point, i. e., a point in whose immediate region an infinite number of points of the sequence are located. Denoting this point of cluster by a, we have here an increasing and a decreasing monotone sequence which both converge towards a, or:

$$\lim_{n=\infty} u_n' = \lim_{n=\infty} u_n = a.$$

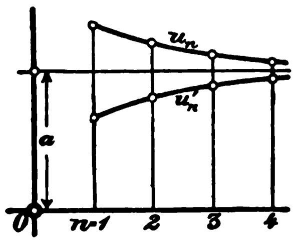
This relation may be illustrated by the accompanying diagram:

If we now let $\lim u_n = \lim\ u_n \cdot e^{-1/12n} = a$, then we shall have

for every finite value of n:

$$u_n \cdot e^{-1/12n} < a < u_n,$$

where $a = u_n \cdot e^{-\theta/12n}$ $(0 < \theta < 1)$.



This gives us finally the following expression for $n!$:

$$n! = a \cdot n^{n+1/2} \cdot e^{-n+\theta/12n}. \tag{III}$$

In this expression we need only determine the unknown coefficient a. The formula of Wallis gives immediately:

$$\lim_{n=\infty} \frac{(2 \cdot 4 \cdot 6 \cdots 2n)^2}{(2n)!\sqrt{2n}} = \lim_{n=\infty} \frac{2^{2n}(n!)^2}{(2n)!\sqrt{2n}} = \sqrt{\pi/2}.$$

Substituting in this latter expression the value for factorials as found in (III) and neglecting the quantity: $\theta/12n$, we have after a few reductions:

$$\lim_{n=\infty} \frac{an}{\sqrt{2n(2n)}} = \sqrt{\pi/2}, \quad \text{or } a = \sqrt{2\pi},$$

from which we easily obtain De Moivre-Stirling's Formula in its final form:

$$n! = \sqrt{2\pi} \cdot n^{n+1/2} \cdot e^{-n}.$$

This remarkable approximation formula gives even for comparatively small values of n surprisingly accurate results. Thus for instance we have:

$$10! = 3,628,800; \; 10^{10}e^{-10}\sqrt{20\pi} = 3,598,699.$$

CHAPTER IX.

LAW OF LARGE NUMBERS. MATHEMATICAL DEDUCTION.

59. Repeated Trials.—Let us consider a general domain of action wherein the determining causes remain constant and produce either one or the other of the opposite and mutually exclusive events, E and $\overline{E}$, with the respective a priori probabilities p and q ($q = 1 - p$) in a single trial. The trial (observation) will, however, be repeated s times with the explicit assumption that the outward conditions influencing the different trials remain unaltered during each observation. The simplest example of observations of this kind is offered by repeated drawings of balls from an urn containing white and black balls only, and where the ball is put back in the urn and mixed thoroughly with the rest before the next drawing takes place. We keep now a record of the repetitions of the opposite events, E and $\overline{E}$ during the s trials, irrespective of the order in which these two events may happen. This record must necessarily be of one of the following forms:

E	happens	s	times,	$\overline{E}$ 0	times,
E	"	$s-1$	"	$\overline{E}$ 1	"
E	"	$s-2$	"	$\overline{E}$ 2	"
.	.	.	.	.	.
.	.	.	.	.	.
E	"	0	"	$\overline{E}$ s	"

In Chapter IV, Example 17, we showed that the probabilities of the above combinations of the two events, E and $\overline{E}$, were determined by the expansion of the binomial

$$(p + q)^s.$$

The general term

$$\frac{s!}{\alpha!\,\beta!}p^\alpha q^\beta \quad (\alpha + \beta = s)$$

is the probability $P(E^\alpha \bar{E}^\beta)$ that E will happen α and $\bar{E}$ β times in the s total trials. Each separate term of the binomial expansion of $(p + q)^s$, represents the probability of the happening of the two events in the order given in the above scheme.

60. Most Probable Value.—In dealing with these various terms, it has usually been the custom of the English and French mathematicians as well as many German scholars to pay particular attention to a special term, the maximum term, which generally is known as the "most probable value" or the "mode." Russian and Scandinavian writers and the followers of the Lexis statistical school of Germany have preferred to make another quantity known as the "probable" or "expected value," the nucleus of their investigations. Although it is our intention to follow the latter method, we shall discuss first, briefly, the most probable value. Two questions are then of special interest to us:

(1) What particular event is most probable to happen?

(2) What is the probability that an event will occur whose probability does not differ from that of the most probable event by more than a previously fixed quantity?

Neither of the two questions offers any particular principal difficulties from a theoretical point of view. When regarding the probability $P(E^\alpha \bar{E}^\beta)$, which we shall denote by T, as a function of the variable quantity, α, T evidently will reach a maximum value for a certain value of α, $(\beta = s - \alpha)$, and we need only determine the greatest term in the above binomial expansion.

In order to answer the second question we have only to pick out all the terms which are situated between the two fixed limits. Their sum is then the probability that those two limits are not exceeded.

61. Simple Numerical Examples.—When s is a comparatively small number the actual expansion may be performed by simple arithmetic. We shall, for the benefit of the student, give simple example of this kind.

A pair of dice is thrown 4 times in succession, to investigate the chance of throwing doublets.

In a single throw the probability of getting a doublet is $p = \frac{1}{6}; \left(q = \frac{5}{6}\right)$. Expanding $\left(\frac{1}{6} + \frac{5}{6}\right)^4$ by means of the binominal theorem we get $\left(\frac{1}{6}\right)^4 + 4\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right) + 6\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 + 4\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^3 + \left(\frac{5}{6}\right)^4$. Each of the above terms represents the probability of the occurrence of the various combinations of doublets (E) and non-doublets ($\bar{E}$), and it is readily seen that the event of getting no doublets at all, represented by the last term $\left(\frac{5}{6}\right)^4 = 0.4823$, has the greatest probability. In other words it is the *most probable* event.

Let us next repeat the trial 12 times instead of 4. The 13 possible probabilities for the various combinations of doublets and non-doublets will then be expressed by the respective terms in the expression

$$\left(\frac{1}{6} + \frac{5}{6}\right)^{12}.$$

The 13 members have as their common denominator the quantity 2,176,782,336 and as numerators the following quantities: 1, 60, 1,650, 27,500, 309,375, 2,475,000, 14,437,500, 61,875,000, 193,359,375, 429,687,500, 644,531,250, 585,937,500, 244,140,625, which now shows that the most probable combination is the one of 2 doublets and 10 non-doublets, having a numerical value equal to .2961.

A further comparison will show that the most probable event in the second series had the probability .2961, whereas .4823 was its value in the first series. In other words, the probability decreases when the trials (observations) are increased. This is due to the fact that the total number of possible cases becomes large with the increase of experiments.

Another question which presents itself, in this connection, is the following: What is the probability that an event will occur

whose probability does not differ from the most probable value by more than a previously fixed quantity? Let us suppose we were asked to determine the probability that a doublet does not occur oftener than 5 times and not less than 1 time in 12 trials. This probability is found by adding the numerical values of the probabilities as given in the binomial expansion from the term containing $p = \frac{1}{6}$ to the power 6, to p to the first power or

$$\frac{14{,}437{,}500 + 6{,}187{,}500 + 193{,}359{,}375 + 429{,}687{,}500 + 644{,}531{,}250 + 585{,}937{,}500}{2{,}176{,}782{,}336}$$

62. The Most Probable Value in a Series of Repeated Trials.—In the examples just given we determined the probability for the happening of the most probable event in a series of s observations by a direct expansion of the binomial $(p + q)^s$. This may be done whenever s is a comparatively small number. But, when s takes on large values, this method becomes impracticable, not to say impossible. Suppose that $s = 1{,}400$, then the actual straightforward expansion $(p + q)^{1400}$ would require a tremendous work of calculation which no practical computer would be willing to undertake. We must therefore in some way or other seek a method of approximation by which this labor of calculation may be avoided and try to find an approximate formula by which we are able to express the maximum term in a simple manner, involving little computation and at the same time yielding results close enough for practical as well as theoretical purposes. Jacob Bernoulli in his famous treatise "Ars Conjectandi" was the first mathematician to solve this problem. Bernoulli also gave an expression for the probability that the departure from the most probable value should not exceed previously fixed limits. The method, however, was very laborious and the final form was first reached by Laplace in "Théorie des Probabilités."

We saw before in Chapter IV that the general term

$$T = \frac{s!}{\alpha!\,\beta!}\,p^\alpha q^\beta \quad (\alpha + \beta = s)$$

in the binomial expansion $(p + q)^s$ represented the probability that an event, E, will happen α times and fail β times in s trials, where p and q were the respective probabilities for success and failure in a single trial. The exponent α may here take all positive integral values in the interval $(0, s)$, including both limits. The question now arises, which particular value of α, say α_n, will make the above quantity a maximum term in the expansion of the binomial? If α_n really is this particular value, then it must satisfy the following inequalities:

$$\underset{\text{(I)}}{\frac{s!}{(\alpha_n + 1)!(\beta_n - 1)} p^{\alpha_n+1}q^{\beta_n-1}} \leqq \underset{\text{(II)}}{\frac{s!}{\alpha_n!\beta_n!} p^{\alpha_n}q^{\beta_n}}$$

$$\geqq \underset{\text{(III)}}{\frac{s!}{(\alpha_n - 1)!(\beta_n + 1)!} p^{\alpha_n-1}q^{\beta_n+1}}.$$

Dividing (II) by (III) and (II) by (I) we obtain the following inequalities:

$$\frac{\beta_n + 1}{\alpha_n}\frac{p}{q} \geqq 1 \text{ and } \frac{\alpha_n + 1}{\beta_n}\frac{q}{p} \geqq 1,$$

which also may be written

$$(\beta_n + 1)p \geqq q\alpha_n \text{ and } (\alpha_n + 1)q \geqq \beta_n p.$$

The following reductions are self evident:

$$(s - \alpha_n + 1)p \geqq \alpha_n(1 - p) \text{ or } sp + p \geqq \alpha_n,$$

and

$$(\alpha_n + 1)q \geqq (s - \alpha_n)p \text{ or } \alpha_n q + \alpha_n p \geqq sp - q \text{ or } \alpha_n \geqq sp - q.$$

From which we see that α_n satisfies the following relation:

$$ps - q \leqq \alpha_n \leqq ps + p.$$

Since $p + q = 1$, we notice that α_n is enclosed between two limits whose difference in absolute magnitude equals unity. The whole interval in which α_n is situated being equal to unity, and since α_n must be an integral number, this particular α_n is determined uniquely as an integral positive number when both $ps - q$ and $ps + p$ are fractional quantities. If $ps - q$ is an integral number $ps + p$ will also be integral, and α_n had to be a

fractional number in order to satisfy the above inequality. Since by the nature of the problem α_n can take positive integral values only, the binomial expansion of $(p + q)^s$ must have two terms which are greater than any of the rest. Dividing both sides of the inequality by s, we shall have

$$p - \frac{q}{s} < \frac{\alpha_n}{s} < p + \frac{p}{s} \text{ or } q + \frac{q}{s} > \frac{\beta_n}{s} \text{ and } p + \frac{p}{s} > \frac{\alpha_n}{s}.$$

Since both p and q are proper fractions, both p/s and q/s are less than $1/s$. We may therefore safely assume that the highest possible difference between the two quotients α_n/s and β_n/s and the probabilities p and q will never exceed $1/s$. Now if s is a very large number this quantity may be neglected, and we may therefore write $ps = \alpha_n$ and $qs = \beta_n$.

Substituting these values in our original expression for the general term of the binomial expansion we get as the maximum number:

$$T_m = \frac{s!}{(sp)!(sq)!} p^{sp} q^{sq}.$$

63. Approximate Calculation of the Maximum Term, T_m.— When the trials are repeated a large number of times the straightforward calculation of the maximum term becomes very laborious. The only table facilitating an exact computation is in a work "Tabularum ad Faciliorem et Breviorem Probabilitatis Computationem Utilium, Enneas," by the Danish mathematician, C. F. Degen. This table, which was published in 1824, gives the logarithms to twelve places for all values of $n!$ from $n = 1$ to $n = 1{,}200$. Degen's table is, however, not easily obtained, and even if it were, it would be of little or no value for factorials above $1{,}200!$. Our only resort is therefore to find an approximate expression for the above value of $n!$. This is most conveniently done by making use of Stirling's formula for factorials of high orders. We have

$$s! = s^{s+1/2} e^{-s} \sqrt{2\pi},$$

$$(sp)! = (sp)^{sp+1/2} e^{-sp} \sqrt{2\pi},$$

$$(sq)! = (sq)^{sq+1/2} e^{-sq} \sqrt{2\pi}.$$

Substituting the above values in the expression $s!/((sp)!\,(sq)!)$ we get

$$\frac{1}{p^{sp+1/2}q^{sq+1/2}\sqrt{2\pi s}}.$$

Hence we have

$$T_m = \frac{p^{sp}q^{sq}}{p^{sp+1/2}q^{sq+1/2}\sqrt{2\pi s}},$$

which reduces to

$$T_m = \frac{1}{\sqrt{2\pi spq}}$$

as an approximate value for the maximum term.

Tchebycheff's Theorems.—Despite all that has been said about the most probable value, its use is somewhat limited, and it might well, without harm, be left out of the whole theory of probabilities. Just because an event is the most probable it does by no means follow it is a very probable event. In fact the expression $(\sqrt{2\pi spq})^{-1}$ which for large values of s converges towards zero shows that the most probable event in reality is a very improbable event. This statement may seem a little paradoxical; but it is easily understood by realizing that the most probable event is only a probability for a possible combination among a large number of equally possible combinations of a different order.

Instead of finding the most probable event it is more important in practical calculations to determine the average number or mean value of the absolute frequencies of successes. In Chapter V we pointed out the close relation between a mathematical expectation and the mean value of a variable. This relation is used by the Russian mathematician, Tchebycheff, as the basis of some very general and far-reaching theorems in probabilities, by means of which the Law of Large Numbers may be established in an elegant and elementary manner.

64. Expected or Probable Value.—In Chapter V we defined the product of a certain sum, s, and the probability of winning such a sum as the mathematical expectation of s. It is, however, not necessary to associate the happening of the event with a monetary gain or loss, in fact it serves often to confuse the reader and we may generalize the definition as follows. *If a*

variable α_i may assume any of the values $\alpha_1, \alpha_2, \alpha_3 \cdots \alpha_s$ each with a respective probability of existence $\varphi(\alpha_i)$ $(i = 1, 2 \cdots s)$ and such that $\Sigma\varphi(\alpha_i) = 1$, then we define:

$$\Sigma\alpha_i\varphi(\alpha_i) = e(\alpha_i)$$

as the expected value of α_i.

Some writers use also the term probable value instead of expected value. In other words the expected value of a variable quantity, α, which may assume any one of the values $\alpha_1, \alpha_2 \cdots \alpha_s$ is the sum of the products of each individual value of the variable and the corresponding probability of existence of such value.

Suppose we now have two opposite and complementary events E and $\bar{E}$ for which the probabilities of happening in a single trial are equal to p and $q = 1 - p$ respectively. When the trials are repeated s times the probabilities of E happening s times, $\bar{E}$ no times, of E happening $s - 1$ and $\bar{E}$ once, of E $s - 2$ and $\bar{E}$ 2 times and so on, may be expressed by the individual terms of the expansion:

$$(p + q)^s,$$

where the general term expressing the occurrence of E α times and of $\bar{E}$ $(s - \alpha)$ times is:

$$\varphi(\alpha) = \binom{s}{\alpha} p^\alpha q^{s-\alpha},$$

which is also the probability of the existence of the frequency number α. The variable in the binomial expansion is α, which may assume all values from 0 to s inclusive.

We now first of all proceed to find the expected value—or the mathematical expectation—of the following quantities:

$$\alpha, \ [\alpha - e(\alpha)] \text{ and } [\alpha - e(\alpha)]^2.$$

We shall presently show the reason for the selection of the above expressions, which perhaps may appear at the present, somewhat puzzling to the student.

In mathematical symbols the expected values of the above quantities are expressed as follows·

$$e(\alpha) = \Sigma\alpha\varphi(\alpha), \ e[\alpha - e(\alpha)] = \Sigma[\alpha - e(\alpha)]\varphi(\alpha)$$

and

$$e[\alpha - e(\alpha)]^2 = \Sigma[\alpha - e(\alpha)]^2\varphi(\alpha)$$

and the summation is to take place from $\alpha = 0$ and to $\alpha = s$.

65. Summation Method of Laplace. The Mean Error.—The analytical difficulty lies in the summation of the expressions as given above. Laplace was the first to give a compact expression for the different sums in a simple and elegant manner. By the introduction of the parameter t Laplace writes:

$$\varphi(\alpha) = (p + q)^s = \Sigma \binom{s}{\alpha} p^\alpha q^{s-\alpha}$$

as

$$\varphi(t\alpha) = (tp + q)^s = \Sigma \binom{s}{\alpha} (tp)^\alpha q^{s-\alpha}.$$

Differentiating with respect to t, which it must be remembered is introduced as an auxiliary parameter only, we have:

$$\varphi'(t\alpha) = sp(tp + q)^{s-1} = \Sigma \alpha p \binom{s}{\alpha} (tp)^{\alpha-1} q^{s-\alpha}.$$

Letting t assume the special value 1 the above sum becomes $e\ (\alpha)$ or

$$e(\alpha) = \Sigma \alpha \binom{s}{\alpha} p^\alpha q^{s-\alpha} = sp(p + q)^{s-1} = sp. \qquad \text{(I.)}$$

We might, however, have obtained the same result in a much shorter manner by the following consideration. The expectation for a single event among the s events is equal to p. Since all the events are independent of each other, it follows from the addition theorem that the complete expectation of the total s cases is equal to sp.

We next proceed to determine the expression: $e[\alpha - e(\alpha)]$ or the expected value of the differences between the constant, $e(\alpha) = sp$ and the individual values 1, 2, 3, $\cdots$, s which α may assume in the binomial expansion.

The difference $\alpha - e(\alpha)$ is known as the departure or deviation from the expected value, some of these deviations will be positive, namely all the values situated to the right of the maximum term, which also is the most probable term in the expansion $(p + q)^s$, while the α's situated to the left of the maximum value of α will be less in magnitude than the largest $\alpha = sp$ and the deviation will therefore be negative. On account of the symmetrical form of the binomial expansion we may expect an

equal number of positive and negative deviations which, taken two and two at a time, are equal in absolute magnitude. The algebraic sum of all the deviations may therefore be expected to be equal to zero. We shall, however, in a rigidly analytical manner prove that this is actually so. We have

$$e[\alpha - e(\alpha)] = \Sigma[\alpha - e(\alpha)]\varphi(\alpha) = \Sigma\alpha\varphi(\alpha) - \Sigma e(\alpha)\varphi(\alpha)$$
$$= \Sigma\alpha\varphi(\alpha) - sp\Sigma\varphi(\alpha).$$

The first term in this last expression we found, however, to be equal to $e(\alpha) = sp$, and we have finally:

$$e[\alpha - e(\alpha)] = sp - sp = 0.$$

By squaring the quantity, $\alpha - e(\alpha)$, we get $\alpha^2 - 2\alpha e(\alpha) + [e(\alpha)]^2$, which is always positive no matter if the above difference is negative.

As a preliminary step we shall find

$$e(\alpha^2) = \Sigma\alpha^2\varphi(\alpha).$$

Introducing the auxiliary parameter, t, we get:

$$\Sigma\binom{s}{\alpha}(tp)^\alpha q^{s-\alpha} = (tp + q)^s.$$

The first derivative with respect to t is:

$$\Sigma p\alpha\binom{s}{\alpha}(tp)^{\alpha-1}q^{s-\alpha} = sp(tp + q)^{s-1}.$$

Multiplying both sides of the equation by tp, we have:

$$\Sigma p\alpha(tp)^\alpha q^{s-\alpha}\binom{s}{\alpha} = stp^2(tp + q)^{s-1}.$$

Differentiating we get:

$$\Sigma p^2\alpha^2\binom{s}{\alpha}(tp)^{\alpha-1}q^{s-\alpha} = sp^2(tp + q)^{s-1} + s(s - 1)p^3t(tp + q)^{s-2}.$$

Dividing through with the constant factor p and letting $t = 1$ we have:

$$\Sigma\alpha^2\binom{s}{\alpha}p^\alpha q^{s-\alpha} = s^2p^2 + sp(1 - p) = s^2p^2 + spq.$$

The expression on the left side is, however, nothing less than the algebraic sum of $\Sigma\alpha^2\varphi(\alpha)$ or simply $e(\alpha^2)$. This leaves the final result:

$$e(\alpha^2) = s^2p^2 + spq.$$

We have now:

$$[\alpha - e(\alpha)]^2 = \alpha^2 - 2\alpha e(\alpha) + [e(\alpha)]^2,$$

from which it follows:

$$e[\alpha - e(\alpha)]^2 = s^2p^2 + spq - 2s^2p^2 + s^2p^2 = spq.$$

Denoting this latter quantity by the symbol $[\epsilon(\alpha)]^2$ we have:

$$[\epsilon(\alpha)]^2 = e[\alpha - e(\alpha)]^2 = spq, \text{ or } \epsilon(\alpha) = \sqrt{spq}. \qquad \text{(II.)}$$

The quantity $\epsilon(\alpha)$ or simply ϵ is commonly known as the mean error of the frequency number α in the Bernoullian expansion. The mean error is one of the most useful functions in the theory of probabilities and furnishes one of the most powerful tools of the statistician.

66. Mean Error of Various Algebraic Expressions.—We next proceed to prove some general theorems connected with the mean error. The mean error of the sum of two observed variables, α and β, is given by the formula:

$$\epsilon(\alpha + \beta) = \sqrt{\epsilon^2(\alpha) + \epsilon^2(\beta)}.$$

Proof: Let $e(\alpha) = \Sigma\alpha\varphi(\alpha)$ and $e(\beta) = \Sigma\beta\psi(\beta)$

$$\epsilon^2(\alpha) = \Sigma[\alpha - e(\alpha)]^2\varphi(\alpha) \text{ and } \epsilon^2(\beta) = \Sigma[\beta - e(\beta)]^2\psi(\beta)$$

be the respective expressions for the probable values and the mean errors of α and β where of course $\Sigma\varphi(\alpha) = 1$ and $\Sigma\psi(\beta) = 1$. Now $\varphi(\alpha_\nu)$ is the probability for the occurrence of the special value α_ν of the variable values, in the same way $\psi(\beta_\mu)$ is the probability for the occurrence of β_μ. If α and β are independent of each other, then according to the multiplication theorem, $\varphi(\alpha_\nu)\psi(\beta_\mu)$ represents the probability for the simultaneous occurrence of α_ν and β_μ as well as the probability of the occurrence of the difference: $\alpha_\nu + \beta_\mu - e(\alpha) - e(\beta)$, since the probable values $e(\alpha)$ and $e(\beta)$ are constant quantities independent of either α or β.

If ϵ denotes the mean error of $\alpha + \beta$ then it follows from the definition of ϵ that $\epsilon^2 = \Sigma\Sigma[\alpha + \beta - e(\alpha) - e(\beta)]^2\varphi(\alpha)\psi(\beta)$ where the double summation is to take place for all possible values of the variables α and β.

The above expression may be written as:

$$\epsilon^2 = \Sigma\Sigma[\alpha - e(\alpha) + \beta - e(\beta)]^2\varphi(\alpha)\psi(\beta),$$

or

$$\epsilon^2 = \Sigma\Sigma[\alpha - e(\alpha)]^2\varphi(\alpha)\psi(\beta) + 2\Sigma\Sigma[\alpha - e(\alpha)][\beta - e(\beta)]\varphi(\alpha)\psi(\beta) + \Sigma\Sigma[\beta - e(\beta)]^2\varphi(\beta\alpha)\psi(\beta).$$

A mere inspection will satisfy that the first and the last terms of this expression equals $\epsilon^2(\alpha)$ and $\epsilon^2(\beta)$ respectively. The first term may be written as follows:

$$\Sigma[\alpha - e(\alpha)]^2\varphi(\alpha)\Sigma\psi(\beta) = \epsilon^2(\alpha)$$

since $\Sigma\psi(\beta) = 1$. The same also holds true for the last term. With regard to the middle term we found before that

$$e[\alpha - e(\alpha)] = 0.$$

Hence it follows by mere inspection that this term becomes 0. Thus we finally have:

$$\epsilon^2(\alpha + \beta) = \epsilon^2(\alpha) + \epsilon^2(\beta) \text{ or } \epsilon(\alpha + \beta) = \sqrt{\epsilon^2(\beta) + \epsilon^2(\alpha)}.$$

Since the middle term is always 0, it follows a fortiori

$$\epsilon(\alpha - \beta) = \sqrt{\epsilon^2(\alpha) + \epsilon^2(\beta)},$$

also that

$$\epsilon(k\alpha) = k\epsilon(\alpha),$$

where k is a constant. This gives us the following theorems: The mean error of the sum or of the difference of two quantities is equal to the square root of the sum of the squares of each separate mean error. The mean error of any quantity multiplied by a constant is equal to this same constant multiplied by the mean error of the quantity. (See Appendix.)

The above theorems may easily be extended to any number of variables: $\alpha, \beta, \gamma \cdots$ so that in general we have

$$\epsilon(\alpha + \beta + \gamma \cdots) = \sqrt{\epsilon^2(\alpha) + \epsilon^2(\beta) + \epsilon^2(\gamma) + \cdots}.$$

We shall later make use of this formula by a comparison of the different rates of mortality among different population groups.

So far we have computed the mean error for the absolute frequencies of α, and the quantity $\sqrt{spq}$ was compared with the most probable number of successes sp. But it may also be useful to know the mean error of the relative frequencies. This calculation is performed by reducing the mean error of the absolute

frequencies to the same degree as these absolute frequencies are reduced to relative frequencies. We saw before that $e(\alpha) = sp$. The relative frequency of the probable value is $e(\alpha)/s = sp/s = p$. The mean error of p therefore is

$$\epsilon[e(\alpha):s] = \sqrt{\frac{pq}{s}}.$$

The following remarks of Westergaard are worthy of note: "When a length is measured in meters and this measure may be effected with an uncertainty of say 2 meters, the length in centimetres is then simply found by multiplication by 100 and the uncertainty is 200 cm. When we wish to find the mean error of p instead of sp we only need to divide the mean error $\sqrt{spq}$ by s, which gives $\sqrt{pq/s}$."

The same result is also easily obtained from the formula

$$\epsilon(k\alpha) = k\epsilon(\alpha)$$

when we let $k = 1/s$.

67. Tchebycheff's Theorem.—Tchebycheff's brochure appeared first in Liouville's Journal for 1866 under the title "Des valeurs Moyennes." A later demonstration was given by the Italian mathematician, Pizetti, in the annals of the University of Geneva for 1892. The nucleus in both Tchebycheff's and Pizetti's investigations is the expression for the mean error:

$$\epsilon(\xi) = \Sigma[\xi - e(\xi)]^2\varphi(\xi). \tag{1}$$

The variable ξ may be of any form whatsoever, it may thus for instance be the sum of several variables: $\alpha, \beta, \gamma \cdots$ while $\varphi(\xi)$ is the ordinary probability function for the occurrence of ξ. Let us denote the difference: $\xi_r - e(\xi_r)$ by $v_r(r = 1, 2, 3 \cdots s)$. We may then write the above expression for $\epsilon(\xi)$ as:

$$\varphi(\xi_1)\frac{v_1^2}{a^2} + \varphi(\xi_2)\frac{v_2^2}{a^2} + \varphi(\xi_3)\frac{v_3^2}{a^2} + \cdots \varphi(\xi_s)\frac{v_s^2}{a^2} = \frac{\epsilon^2(\xi)}{a^2} \tag{2}$$

where a is an arbitrarily chosen constant, but always larger than $\epsilon(\xi)$ in absolute magnitude. If we, in the above equation, select all the v's which are larger than a in absolute magnitude together with their corresponding probabilities, $\varphi(\xi)$ and denote all

such quantities by v', v'', v''', $\cdots$ and $\varphi(\xi)'$, $\varphi(\xi)''$, $\varphi(\xi)'''$, $\cdots$ respectively, we have evidently:

$$\frac{\varphi(\xi)'v'^2}{a^2} + \frac{\varphi(\xi)''v''^2}{a^2} + \frac{\varphi(\xi)'''v'''^2}{a^2} + \cdots < \frac{\epsilon^2(\xi)}{a^2} \tag{3}$$

For any one of these different v's which is larger in absolute magnitude than a

$$\frac{v^2}{a^2} > 1$$

from which it follows a fortiori:

$$\varphi(\xi)' + \varphi(\xi)'' + \cdots = \Sigma\varphi^i(\xi) < \frac{\epsilon^2(\xi)}{a^2}. \tag{3a}$$

In this latter inequality, $\Sigma\varphi^i(\xi)$ is the total probability for the occurrence of a deviation from $e(\xi)$ larger than a in absolute magnitude.

Let now P_T be the probability that the absolute value of the mean error *is not* larger than a; then $1 - P_T$ is the total probability that the mean error *is* larger than a. We have thus from the inequality (3a)

$$1 - P_T < \frac{\epsilon^2(\xi)}{a^2} \quad \text{or} \quad P_T > 1 - \frac{\epsilon^2(\xi)}{a^2}. \tag{4}$$

Let also $a = \lambda\epsilon(\xi)$. We then have by a mere substitution in the above inequality:

$$P_T > 1 - \frac{1}{\lambda^2}. \tag{5}$$

This constitutes the first of Tchebycheff's criterions which says:

The probability that the absolute value of the difference $| \alpha - e(\alpha) |$ *does not exceed the mean error by a certain multiple,* λ, ($\lambda > 1$) *is greater than* $1 - (1/\lambda^2)$.

Now we made no restrictions as to the variable, ξ, which may be composed of the sum of several independent variables, α, β, γ, $\cdots$. We saw before that

$$\epsilon^2(\alpha + \beta + \gamma + \cdots) = \epsilon^2(\alpha) + \epsilon^2(\beta) + \epsilon^2(\gamma) + \cdots$$

Tchebycheff's criterion may therefore be extended as follows:

The Tchebycheffian probability, P_T, *that the difference* $| \alpha + \beta + \gamma + \cdots - e(\alpha) - e(\beta) - e(\gamma) - \cdots |$ *will never exceed the mean error* ϵ *by a certain multiple,* $\lambda > 1$, *is greater than* $1 - (1/\lambda^2)$.

68. The Theorems of Poisson and Bernoulli proved by the application of the Tchebycheffian Criterion.—Bernoulli in his researches limited himself to the solution of the problem in which the probabilities for the observed event remained constant during the total number of observations or trials. Poisson has treated the more general case, wherein the individual probability for the happening of the event in a single trial varies during the total s trials. This may probably best be illustrated by an urn schema. Suppose we have s urns $U_1, U_2, \cdots U_s$ with white and black balls in various numbers. Let the probability for drawing a white ball from the urns $U_1, U_2, \cdots U_s$ in a single trial be $p_1, p_2, \cdots p_s$ respectively, $q_1, q_2, \cdots q_s$ the chances for drawing a black ball in a single trial. If a ball is drawn from each urn, what is the probability of a drawing α white and $s - \alpha$ black balls in s trials? It is easily seen that the Bernoullian Theorem is a special case when the contents of the s urns and the respective probabilities for drawing a white ball in a single trial are the same for all urns.

69. Bernoullian Scheme.—We shall now show how the Tchebycheffian critierions may be used in answering the question given above. First of all we shall start with the simpler case of the Bernoullian urn-schema. Here the probability for drawing a white or a black ball from each of the s urns in a single trial is p and q respectively. The square of the mean error in a single trial is pq. From the formulas in § 66 it then follows:

$$\epsilon^2 = \epsilon_1^2 + \epsilon_2^2 + \cdots = pq + pq + pq + \cdots s \text{ times} = spq$$

or

$$\epsilon = \sqrt{spq}.$$

While the above expression gives us the mean error of the absolute frequency of the variable α, the relative frequency of α to the total number of trials, s, is given as

$$\epsilon = \frac{\sqrt{pq}}{\sqrt{s}}.$$

We now ask: What is the total probability that the absolute deviation of the relative frequency α/s from its expected value $sp/s = p$ never becomes larger than λ times the mean error,

$\epsilon = \sqrt{pq/s}$? Letting $\lambda = \sqrt{s}/t$ and using the symbols P_T for this particular probability, we have according to Tchebycheff's criterion:

$$P_T > 1 - 1/\lambda^2, \quad \text{or} \quad P_T > 1 - t^2/s.$$

Since the mean error is equal to $\sqrt{pq/s}$ we have:

$$\lambda\epsilon = \frac{\sqrt{pq}}{t}.$$

The answer to our question above follows now a fortiori as follows:

The total probability that the absolute deviation of the relative frequency from the postulated a priori probability, p, never exceeds the quantity, $\sqrt{pq}/t$, is greater than $1 - (t^2/s)$.

By taking t large enough we may reduce $\sqrt{pq}/t$ (where pq is a fraction whose maximum value never can exceed $1 \div 4$,) below any previously assigned quantity, δ, however small. If, for instance, we choose the value .0001 for δ, we may rest assured that $\sqrt{pq}/t$ will be less than δ when we take t larger than 5000. But no matter how large t is, so long as it remains a finite number, by letting $s = \infty$ as a limiting value, t^2/s will simultaneously approach 0 as a limiting value. From the deductions thus derived we are now able to draw the following conclusions:

1) *By letting $s = \infty$ as a limiting value, the probability, P_T, that the absolute difference between the relative frequency α/s and the postulated a priori probability, p, never becomes greater than $\sqrt{pq}/t$ approaches 1 or certainty as a limit.*

2) *By choosing the quantity, t, which is less than $\lim_{s \doteq \infty} \sqrt{s}$, sufficiently great, we may bring $\sqrt{pq}/t$ below any previously assigned quantity, δ, or make the difference between p and α/s as small as we please.*

From these conclusions we obtain a fortiori the following

$$\lim_{s=\infty} \frac{\alpha}{s} = p.$$

This constitutes the essential features of the Bernoullian Theorem.

70. Poisson's Scheme.—Let p_1 denote the postulated probability for success in the first trial, p_2 in the second, p_3 in the

third, etc., and let furthermore $q_1, q_2, q_3, \cdots$ be the respective probabilities for the corresponding failures. If the trial (observation) is repeated s times we obtain the following values for the probable or expected value of the frequency for successes $e(\alpha)$ and the mean error ϵ

$$e(\alpha) = p_1 + p_2 + p_3 + \cdots p_s = \Sigma p_i,$$

$$\epsilon = \sqrt{p_1q_1 + p_2q_2 + p_3q_3 + \cdots p_sq_s} = \sqrt{\Sigma p_iq_i} \quad (i = 1, 2, 3, \cdots s)$$

If by p_0 and q_0 we denote the arithmetic mean or the average value of the s p's and s q's, such that

$$p_0 = \frac{p_1 + p_2 + p_3 + \cdots p_s}{s} \tag{3}$$

$$q_0 = \frac{q_1 + q_2 + q_3 + \cdots q_s}{s} \tag{4}$$

and assume that p_0 and q_0 denote the constant probabilities during each of the s trials (observations), we should according to the Bernoullian Theorem have:

$$e(\alpha_B) = sp_0 \tag{5}$$

$$\epsilon(\alpha_B) = \sqrt{sp_0q_0} \tag{6}$$

where α_B stands for the absolute frequency in a Bernoullian series.

An actual comparison of (1) and (5) and (3) shows that:

$$e(\alpha_P) = e(\alpha_B) \tag{7}$$

where α_P is the symbol for the absolute frequency in a Poisson series. In other words: If the s trials had been performed with constant probability for success equal to p_0 instead of with varying probabilities $p_1, p_2, \cdots p_s$, the expected or probable value would be the same for the Bernoullian and Poisson scheme.

With regard to the mean error we find, however, after a little calculation,

$$\epsilon_P^2(\alpha) = \epsilon_B^2(\alpha) - \Sigma(p_i - p_0)^2. \tag{8}$$

The expression for the mean error in Poisson's Theorem is of the following form

$$\epsilon_P = \sqrt{p_1q_1 + p_2q_2 + p_3q_3 + \cdots p_iq_i} = \sqrt{\Sigma p_iq_i} \quad (i = 1, 2, 3 \cdots s).$$

Now $p_i q_i$ may be transformed as follows:
Writing

$$p_i = p_0 + (p_i - p_0)$$

$$q_i = q_0 - (p_i - p_0)$$

and multiplying we obtain:

$$p_i q_i = p_0 q_0 - (p_i - p_0)(p_0 - q_0) - (p_i - p_0)^2,$$

and summing up for all values of i from $i = 1$ to $i = s$ we have:

$$\epsilon_P^2 = s p_0 q_0 - \Sigma(p_i - p_0)^2 = \epsilon_B^2 - \Sigma(p_i - p_0)^2.$$

As $(p_i - p_0)^2$ always is a positive quantity, it is readily seen that the mean error in a Poisson scheme is always less than the mean error in the corresponding Bernoullian series.

Writing ϵ as follows:

$$\epsilon = \sqrt{p_1 q_1 + p_2 q_2 + \cdots + p_s q_s}$$
$$= \sqrt{s}\sqrt{\frac{p_1 + \cdots + p_s}{s} - \frac{p_1^2 + \cdots + p_s^2}{s}}$$

and letting $\lambda = \sqrt{s}/t$, we have according to Tchebycheff's Theorem the following rule: The probability P_T that the relative frequency remains inside the limits:

$$\frac{p_1 + p_2 + \cdots p_s}{s} \pm \frac{\sqrt{s}}{t}\epsilon\left(\frac{\alpha}{s}\right) = \frac{p_1 + p_2 + \cdots + p_s}{s}$$
$$\pm \frac{1}{t}\sqrt{\frac{p_1 + p_2 + \cdots + p_s}{s} - \frac{p_1^2 + p_2^2 + \cdots + p_s^2}{s}}$$

is greater than $1 - (1/\lambda^2)$ or $1 - (t^2/s)$.

By taking t sufficiently large and by letting s approach infinity as a limiting value the last term in the above difference, namely the average probability, p_0, and λ times the mean error, becomes smaller than any previously assigned quantity, δ, however small, while P_T at the same time will approach 1 as a limit.

From this it now follows:

When an infinite number of trials is made on an event, following the scheme of Poisson, then the expression:

$$\lim_{s \to \infty} \frac{\alpha}{s} = \frac{p_1 + p_2 + \cdots + p_s}{s} = p_0.$$

9

The essential part of Poisson's Theorem is contained in this equation. When $p = p_1 = p_2 = \cdots p_s$ we have a Bernoullian series and obtain:

$$\lim_{s=\infty} \frac{\alpha}{s} = p,$$

which result we already derived above in a direct way.

71. Relation between Empirical Frequency Ratios and Mathematical Probabilities.—In the above limit, α indicates the total number of lucky events while s is the total number of trials, the quotient $\alpha \div s$ then is nothing more than the empirical probability as defined in the preceding paragraphs. Both the Bernoullian and Poisson Theorems show that this empirical probability approaches the postulated a priori probability, p, (or the average probability p_0) as a limiting value.

In this way we have succeeded in extending the theory of probability to other problems than the conventional kind involved in the games of chance or drawings of balls from urns. We do not need to limit our investigations to problems where we are able to determine a priori the probability for the happening of an event in a single trial, but limit ourselves to postulate the existence of such an a priori probability.

A large number of trials or observations is made on a certain event E. This event is now observed to have occurred α times during the s total trials. To illustrate: An urn contains red and white balls, the total number of balls being unknown, a single ball is drawn and its color noted. This ball is replaced and the contents of the urn is mixed. A second drawing is made and the color of the drawn ball noted before the ball is put back in the urn. Let this process be repeated s times, where s is a large number, and furthermore let α be the number of red balls which appeared during the s trials.

The quotient $\alpha \div s$ we now call the empirical or a posteriori probability for the observed event, in this particular case the a posteriori probability for the drawing of a red ball. When $s = \infty$ the Bernoullian Theorem tells us that the empirical probability found in this manner and the postulated a priori probability whose numerical value, however, was unknown before the drawings took place, are identical as far as numerical

magnitude is concerned. As we already observed in the introductory remarks to this chapter it is impossible to perform a certain experiment an infinite number of times, and it is therefore out of the question to determine the limiting and ideal value of the posteriori probability, and we must satisfy ourselves with an approximation by performing a finite number of trials, or let s be a finite number. The quotient $\alpha \div s$ is then the empirical approximate a posteriori probability. We know also that although this quotient is an approximation of the postulated a priori probability only, that by increasing s or what amounts to the same thing, by making a large number of trials, the difference between the approximate empirical probability ratio, $\alpha \div s$, and the a priori probability, p, becomes smaller as the number of trials is increased. But how small is the difference? Or how many times shall we repeat the trials (observations) so that, for practical purposes, we may disregard this difference? It does not suffice to be satisfied with the fact that the difference becomes proportionately smaller the greater we make the number of trials and merely insist that in order to avoid large errors it is only necessary to operate with very large numbers. Immediately the question arises: What constitutes a large number? Is 100 a large number, or is 1,000, 10,000, 100,000 or even a million an answer to this question? As long as this question remains unanswered, it helps but little to poke upon the "law of large numbers," a tendency which unfortunately is too manifest in many statistical researches by amateur statisticians. As long as a definition, much less than a numerical determination of the range of "small numbers" is lacking, little stress ought to be laid on such remarks based in the metaphorical terms of "small" and "large" numbers.

72. Application of the Tchebycheffian Criterion.—It is readily seen that even a rough quantitive determination of the difference between the approximate a posteriori probability and the postulated a priori probability based upon the mere vague statement of "large numbers" is utterly impossible, and it remains to be seen, therefore, if the theory of probability offers us a criterion that might serve as a preliminary test for the above difference. To restate our problem: *If p is the postulated a priori*

probability and $\alpha \div s$ is the empirical probability (a posteriori) or relative frequency of the event, E, what is the probability that the difference, $|(\alpha/s) - p|$ does not exceed a previously assigned quantity? In the mean error and the associated theorem of Tchebycheff we have a simple and easily applied criterion to test this probability.

Tchebycheff's rule states that the probability, P_T, of a deviation of a variable from its probable value, not larger than λ times its mean error, is greater than $1 - (1/\lambda^2)$.

For

$$\lambda = 3 \quad P_T > 1 - \tfrac{1}{9} = 0.888$$

$$\lambda = 4 \quad P_T > 1 - \tfrac{1}{16} = 0.937$$

$$\lambda = 5 \quad P_T > 1 - \tfrac{1}{25} = 0.96.$$

This shows that a deviation from the expected or probable value of the variable equal to 4 or 5 times the mean error possesses a very small probability and such deviations are extremely rare.

Let us for example assume that the observed rate of mortality in a certain population group is equal to .0200. Let furthermore the number exposed to risk equal 10,000. The mean error is then $\left(\frac{.02\times.98}{10,000}\right)^{\frac{1}{2}} = .0014$. If the number of lives exposed to risk was one million instead of 10,000, the mean error would be $\left(\frac{.02\times.98}{1,000,000}\right)^{\frac{1}{2}} = .00014$. A deviation four times this latter quantity is equal to .00056, and according to Tchebycheff's criterion the probability for the *non-occurrence* of a deviation above .00056 is greater than .937, or the probability of dying inside a year will not be higher than .0206 or less than .0194. For an observation series of 4,000,000 homogeneous elements we might by a similar procedure expect to find a rate of mortality between 0.02 + 0.00028 or 0.02 − 0.00028. Thus we notice that the mean error of the relative frequency numbers decreases as the number of observations increases.

CHAPTER X.

THE THEORY OF DISPERSION AND THE CRITERIA OF LEXIS AND CHARLIER.

73. Bernoullian, Poisson and Lexis Series.—In the previous chapter we limited our discussion to single sets consisting of s individual trials and found in the mean error and the criterion of Tchebycheff a measure for the uncertainty with which the relative frequency ratio α/s as well as the absolute frequency α were affected. How will matters now turn out if, instead of a single set, we make N sets of trials? As already mentioned in paragraph 54, in general in N such sets we shall obtain N different values of α, denoting the absolute frequency of the event represented by the sequence

$$\alpha_1, \quad \alpha_2, \quad \alpha_3, \quad \cdots \alpha_N.$$

Our object is now to investigate whether the distribution of the above values of α around a certain norm is subject to some simple mathematical law and if possible to find a measure for such distributions.

In this connection it is of great importance whether the postulated a priori probabilities remain constant or not during the N sample sets. Three cases are of special importance to us.[1]

1. The probability of the happening of the event remains constant during all the N sets. The series as given by the absolute frequencies in each set is known as a *Bernoullian Series.*

2. The same probability varies from trial to trial inside each of N sample sets, the variations being the same from set to set. The series as given by the absolute frequencies is in this case known as a *Poisson Series.*

3. The probability remains constant in any one particular set but varies from set to set. The absolute frequency series as produced in this way is called a *Lexis Series.*

The above definition of these three series may, perhaps, be made clearer by a concrete urn scheme.

[1] The terminology in due to Charlier.

A. Bernoullian Series.—s balls are drawn one at a time from an urn, containing black and white balls in *constant proportion* during all drawings. Such drawings constitute a sample set. Let us in this particular set have obtained say α_1 white and β_1 black balls, where $\alpha_1 + \beta_1 = s$. We make N sets of drawings under the same conditions, keeping a record of white balls drawn in each set. The number sequence thus obtained,

$$\alpha_1, \quad \alpha_2, \quad \alpha_3, \quad \cdots \alpha_N.$$

is a Bernoullian Series.

B. Poisson Series.—s individual urns contain white and black balls, the proportion of white to black varying from urn to urn. A single ball is drawn from each urn and its color noted. In this way we get α_1 white and β_1 black balls constituting a set. The balls thus drawn are replaced in their respective urns and a second set of s drawings is performed as before, resulting in α_2 white and β_2 black balls. The number sequence,

$$\alpha_1, \quad \alpha_2, \quad \alpha_3, \quad \cdots \alpha_N,$$

of white balls in N sets represents a Poisson Series.

C. Lexis Series.—s balls are drawn one at a time under the same conditions as set No. 1 in the Bernoullian series. The α_1 white and β_1 black thus drawn constitute the first set. In the second and following set the composition of the urn is changed from set to set. The number sequence representing the number of white balls in the N respective sets:

$$\alpha_1, \quad \alpha_2, \quad \alpha_3, \quad \cdots \alpha_N$$

is a Lexian Series. The scheme of drawings is the same as in the Bernoullian Series except that the proportion of white to black balls varies from set to set.

74. The Mean and Dispersion.—Since we have no a priori reasons for choosing any one particular value of the various α's of the above sequences in preference to any other, we might give equal weight to each set and take the arithmetic mean as defined by the formula:

$$M = \frac{\alpha_1 + \alpha_2 + \alpha_3 + \cdots \alpha_N}{N} \tag{I}$$

of the N values of α.

It will be unnecessary to enter into a detailed discussion of the mean, which is a quantity used on numerous occasions in every day life. We shall, however, define another important function known as the *dispersion* (standard deviation). The dispersion is denoted by the Greek letter, σ, and is defined by the formula

$$\sigma^2 = \frac{(\alpha_1 - M)^2 + (\alpha_2 - M)^2 + \cdots (\alpha_N - M)^2}{N}. \qquad \text{(II)}$$

We shall now attempt to find the expected value of the mean and the dispersion in the three series. First of all take the Bernoullian Series. Let the constant probability for success in a single trial be p_0. We have then for the various expected values or mathematical expectations of α:

Set No. 1: $\quad e(\alpha_1) = sp_0$

Set No. 2: $\quad e(\alpha_2) = sp_0$

. . .

. . .

Set No. N: $\quad e(\alpha_N) = sp_0$

or:

$$\frac{e(\alpha_1) + e(\alpha_2) + \cdots + e(\alpha_N)}{N} = \frac{\Sigma e(\alpha_\nu)}{N} = \frac{Nsp_0}{N} = sp_0,$$

which shows that the mean in a Bernoullian Series of N sample sets is equal to the expected value of the absolute frequency in a single set.

In regard to the dispersion we have for the various sets:

Set No. 1: $\quad e(\alpha_1 - M)^2 = \epsilon^2(\alpha_1) = sp_0q_0$

Set No. 2: $\quad e(\alpha_2 - M)^2 = \epsilon^2(\alpha_2) = sp_0q_0$

. . . . . .

. . . . . .

Set No. N: $\quad e(\alpha_N - M)^2 = \epsilon^2(\alpha_N) = sp_0q_0$

Summing up and forming the mean we obtain for the expected value of the dispersion in a Bernoullian Series, which we shall denote by the symbol σ_B:

$$\sigma_B{}^2 = \frac{\Sigma\epsilon^2(\alpha_\nu)}{N} = \frac{Nsp_0q_0}{N} = sp_0q_0.$$

This result shows that the dispersion in a Bernoullian Series is equal to the mean error, ϵ, in a single set.

We now proceed to the Poisson Series. Let p_1 be the mathematical probability of the happening of the event in the first trial, p_2 be the probability in the second trial and so on for all trials, and let us furthermore denote the means of the p's and q's by:

$$p_0 = \frac{p_1 + p_2 + p_3 \cdots + p_s}{s}.$$

$$q_0 = \frac{q_1 + q_2 + q_3 \cdots + q_s}{s}.$$

Applying a similar analysis as above we have:

Set No. 1: $e(\alpha_1) = p_1 + p_2 + \cdots + p_s = sp_0$
Set No. 2: $e(\alpha_2) = p_1 + p_2 + \cdots + p_s = sp_0$
. . . .
. . . .
Set No. N: $e(\alpha_N) = p_1 + p_2 + \cdots + p_s = sp_0$

The actual summation of the above values of $e(\alpha)$ gives us the following value of the mean in a Poisson Series:

$$M_P = sp_0.$$

Let us for a moment assume that all the drawings had been performed with a constant probability, p_0. According to the Bernoullian scheme we should then have:

$$M_B = sp_0.$$

An actual comparison shows that $M_B = M_P$. This shows that the same mean result is obtained if we draw s balls from the urns $U_1, U_2, \cdots U_s$ with their corresponding probabilities $p_1, p_2, \cdots p_s$ for drawing a white ball, as would be obtained if we drew all the s balls from a single urn where the composition is such that the ratio of the number of white to that of black balls is as $p_0 : q_0$, where p_0 and q_0 are defined as above.

Let us now see how matters turn out in regard to the dispersion. We have for the N sets:

Set No. 1: $e(\alpha_1 - M)^2 = p_1q_1 + p_2q_2 + \cdots = \Sigma p_\nu q_\nu = \epsilon^2(\alpha_1)$
Set No. 2: $e(\alpha_2 - M)^2 = p_1q_1 + p_2q_2 + \cdots = \Sigma p_\nu q_\nu = \epsilon^2(\alpha_2)$
. . . .
. . . .
Set No. N: $e(\alpha_N - M)^2 = p_1q_1 + p_2q_2 + \cdots = \Sigma p_\nu q_\nu = \epsilon^2(\alpha_N)$

In § 70 we showed, however, that $\Sigma p_\nu q_\nu$ could be expressed as follows:

$$\epsilon_P^2(\alpha) = sp_0q_0 - \Sigma(p_\nu - p_0)^2 = \epsilon_B^2(\alpha) - \Sigma(p_\nu - p_0)^2.$$

A simple straightforward calculation gives us now for the dispersion, σ_P^2,

$$\sigma_P^2 = \sigma_B^2 - \Sigma(p_\nu - p_0)^2,$$

In the corresponding Bernoullian Series with constant probability, p_0, the dispersion is equal to sp_0q_0, which shows that the dispersion in a Poisson Series is less than the corresponding dispersion of the Bernoullian Series.

We finally come to the mean and the dispersion in the Lexian Series which we shall denote by M_L and σ_L respectively. Let us furthermore define the two quantities p_0 and q_0 as follows:

$$p_0 = \frac{p_1 + p_2 + \cdots + p_N}{N},$$

$$q_0 = \frac{q_1 + q_2 + \cdots + q_N}{N}.$$

A computation along similar lines as above gives us first for the mean, M_L:

Set No. 1:	$e(\alpha_1) = sp_1$
Set No. 2:	$e(\alpha_2) = sp_2$
	$\cdot \quad \cdot \quad \cdot$
	$\cdot \quad \cdot \quad \cdot$
Set No. N:	$e(\alpha_N) = sp_N$

Thus we have:

$$M_L = \frac{\Sigma e(\alpha_\nu)}{N} = \frac{\Sigma sp_\nu}{N} = \frac{s[p_1 + p_2 + \cdots p_N]}{N} = sp_0.$$

For the dispersion we have the following expectations:

Set No. 1:	$e(sp_0 - \alpha_1)^2$
Set No. 2:	$e(sp_0 - \alpha_2)^2$
	$\cdot \quad \cdot \quad \cdot$
	$\cdot \quad \cdot \quad \cdot$
Set No. N:	$e(sp_0 - \alpha_N)^2$

The expected value in the νth set is

$$e(sp_0 - \alpha_\nu)^2 = \Sigma(sp_0 - \alpha_\nu)^2\varphi_\nu(\alpha),$$

where $\varphi_\nu(\alpha)$ is the general term in the probability binomial: $(p_\nu + q_\nu)^s = 1$. An analysis along similar lines as in § 65 gives us now:

$$e(sp_0 - \alpha_\nu)^2 = s^2p_0^2 - 2s^2p_0p_\nu + s^2p_\nu^2 + sp_\nu q_\nu$$
$$= sp_\nu q_\nu + s^2(p_\nu - p_0)^2$$

as the expected value of the square of the difference between the mean and the absolute frequency in the νth set. For all N sets we then have

$$\sigma_L^2 = \frac{\Sigma sp_\nu q_\nu}{N} + \frac{s^2}{N}\Sigma(p_\nu - p_0)^2.$$

We have, however, the following identity:

$$\Sigma p_\nu q_\nu = Np_0q_0 - \Sigma(p_\nu - p_0)^2,$$

and hence

$$\sigma_L^2 = \sigma_B^2 + \frac{s^2 - s}{N}\Sigma(p_\nu - p_0)^2.$$

74a. Mean or Average Deviation.—Of quite another character than the standard deviation or dispersion is the so-called mean or average deviation, ϑ, defined by means of the following relation:

$$\vartheta = \frac{|\alpha_1 - M| + |\alpha_2 - M| + |\alpha_3 - M| + \cdots + |\alpha_N - M|}{N},$$

where $|\alpha_\nu - M|$ means the absolute difference between m_ν and M. We shall now proceed to determine the expected value of ϑ on the assumption that the observed data follow the Bernoullian Law. The mean in a Bernoullian series with constant probability p_0 we found before to be equal to sp_0 which was the expected value of α in a single sample set of s trials. The expected value of the absolute difference in the νth set is therefore:

$$e\,|\alpha_\nu - sp_0| = \Sigma\,|\alpha_\nu - sp_0|\,\varphi_\nu(\alpha),$$

where as usual $\varphi_\nu(\alpha)$ is the binomial probability function.

The deviations from sp_0 are partly positive and partly negative. We proved, however, before that

$$e(\alpha_\nu - sp_0) = \Sigma(\alpha_\nu - sp_0)\varphi_\nu(\alpha) = 0.$$

Hence it is readily seen that the algebraic sum of the positive deviations cancel the algebraic sum of the corresponding negative deviations so that $e \mid \alpha_\nu - sp_0 \mid$ equals twice the sum of the positive deviations. Positive deviations occur for values of α greater than sp_0, *i. e.*, for all values which α may assume from s to sp_0 in the binomial expansion: $(p_0 + q_0)^s$.[1] Hence we have (omitting subscripts):

$$e \mid \alpha - sp \mid = 2 \sum_{\alpha=s}^{sp} (\alpha - sp) \binom{s}{\alpha} p^\alpha q^{s-\alpha}$$
$$= 2 \left\{ \sum_{\alpha=s}^{sp} \alpha \binom{s}{\alpha} p^\alpha q^{s-\alpha} - sp \sum_{\alpha=s}^{sp} \binom{s}{\alpha} p^\alpha q^{s-\alpha} \right\}.$$

The second of these sums represents the following function of p and q

$$f(p, q) = p^s + \binom{s}{1} p^{s-1} q + \binom{s}{2} p^{s-2} q^2 + \cdots + \binom{s}{sq} p^{sp} q^{sq}.$$

By partial differentiation in respect to p and by following multiplication by p we have:

$$p \frac{\partial f}{\partial p} = sp^s + (s-1) \binom{s}{1} p^{s-1} q + (s-2) \binom{s}{2} p^{s-2} q^2 + \cdots$$
$$+ sp \binom{s}{sq} p^{sp} q^{sq}$$
$$= \sum_{\alpha=s}^{sp} \alpha \binom{s}{\alpha} p^\alpha q^{s-\alpha}.$$

Hence we may write:

$$e \mid \alpha - sp \mid = 2 \left\{ p \frac{\partial f}{\partial p} - spf \right\}.$$

Furthermore $f(p, q)$ is a homogenous function in respect to p and q of the sth order. We may then apply the following well known Eulerian Theorem from the differential calculus: If $f(p, q)$ is homogenous and has continuous first partial derivatives then

$$sf = p \frac{\partial f}{\partial p} + q \frac{\partial f}{\partial q}.$$

Using this relation we may write:

$$e \mid \alpha - sp \mid = 2 \left\{ p \frac{\partial f}{\partial p} - spf \right\} = 2pq \left\{ \frac{\partial f}{\partial p} - \frac{\partial f}{\partial q} \right\}.$$

[1] Sp_0 is taken to the nearest integer.

The partial derivatives of $f(p, q)$ with respect to p and q are of the form:

$$\frac{\partial f}{\partial p} = sp^{s-1} + s(s-1)p^{s-2} + \cdots + \frac{s(s-1)\cdots(sp+1)}{1\cdot 2\cdot 3\cdots(sq-1)}p^{sp}q^{sq-1}$$
$$+ \frac{s(s-1)\cdots sp}{1\cdot 2\cdot 3\cdots sq}p^{sp-1}q^{sq}$$
$$\frac{\partial f}{\partial q} = sp^{s-1} + s(s-1)p^{s-2} + \cdots + \frac{s(s-1)\cdots(sp+1)}{1\cdot 2\cdot 3\cdots(sq-1)}p^{sp}q^{sq-1}$$

Hence we have:

$$\frac{\partial f}{\partial q} - \frac{\partial f}{\partial q} = \frac{s(s-1)\cdots sp}{1\cdot 2\cdot 3\cdots sq}p^{sp-1}q^{sq} = s\left\{\frac{s}{sp!\,sq!}p^{sp}q^{sq}\right\}.$$

We proved, however, in § 63 that the expression inside the bracket may be written approximately as follows:

$$T_m = \frac{1}{\sqrt{2\pi spq}}.$$

This gives us finally (again using the subscripts):

$$e\,|\,\alpha_\nu - sp_0\,| = 2sp_0q_0T_m = \sqrt{\frac{2sp_0q_0}{\pi}}$$

as the expected value of the absolute deviation in the νth sample set. This same relation evidently holds true for any other of the N sample sets, which finally gives us the following result for ϑ:

$$\vartheta = \sqrt{\frac{2}{\pi}} \cdot \sqrt{sp_0q_0}.$$

The dispersion in a Bernoullian series we found before to be of the form:

$$\sigma_B = \sqrt{sp_0q_0}.$$

Hence we have the following relation between the dispersion and the mean deviation:

$$\sigma_B = \sqrt{\frac{\pi}{2}}\,\vartheta = 1.2533\,\vartheta.$$

75. The Lexian Ratio and the Charlier Coefficient of Disturbancy.—The results given in the last few paragraphs may be embodied under the following captions.

1. *The mean in a Poisson and Lexis Series is the same as the mean in a Bernoullian Series with constant probability of p_0 in a single trial, where p_0 is defined as above.*

2. *The dispersion in a Poisson Series is less than the corresponding dispersion in a Bernoullian Series.*

3. *The dispersion in a Lexis Series is greater than the dispersion in a Bernoullian Series.*

The mean and the dispersion of the Bernoullian Series occupy in this connection a central position and may be used as a standard of comparison with other series. This is the method adopted by Lexis in investigating certain statistical series, and we shall return to it in the following chapter. Lexis determines first in a direct manner the dispersion as defined by formula (II) from the statistical data as given by the number sequence α. This process is known as the direct process (by Lexis called a physical process) and gives a certain dispersion, σ. After this the dispersion is computed by an indirect (combinatorial) process under the assumption that the series follows the Bernoullian distribution. The ratio, $\sigma : \sigma_B$, which Charlier calls the *Lexian Ratio* and denotes by the symbol, L, may now give us an idea about the real nature of the statistical series as represented by the number sequence.

When $L = 1$, the series is by Lexis called a *normal* series.

When $L > 1$, the series is called *hypernormal.*

When $L < 1$, the series is a *subnormal* series.

It is easily seen from the respective formulas that the Poisson Series are subnormal series whereas the Lexian Series are hypernormal. The great majority of statistical series are—as we shall have occasion to see in the following chapter—of a hypernormal kind and correspond thus to the Lexian Series.

In § 74 we found the dispersion in the Lexis series as

$$\sigma_L^2 = \sigma_B^2 + (s^2 - s)\sigma_p^2,$$

where

$$\sigma_p^2 = \frac{\Sigma(p_\nu - p_0)^2}{N}.$$

The quantity, σ_p, is the natural measure of the variations in the chances from the mean or normal probability, p_0. It is

however, dependent on the absolute values of these chances, so that if all chances are changed in the same proportion, σ_p is also changed in the same proportion. Another drawback which influences the Lexian Ratio is the variations of the number s in each sample set. In order to overcome this difficulty Charlier divides the above quantity σ_p by p_0. Assuming that the variations in the individual probabilities within each set are of no perceptible influence on the dispersion, we have from the Lexian dispersion:

$$\frac{\sigma_p^2}{p_0^2} = \frac{\sigma^2 - \sigma_B^2}{(s^2 - s)p_0^2}.$$

Neglecting s in comparison with s^2 and remembering that $M_B = sp_0$, we have as an approximation:

$$\frac{\sigma_p}{p_0} = \frac{\sqrt{\sigma^2 - \sigma_B^2}}{M_B} = \rho.$$

Charlier calls the quantity 100ρ the *coefficient of disturbancy* of the statistical series. It is readily seen that the Charlier coefficient is zero in normal series. For hypernormal series it is a positive real quantity whereas for subnormal series ρ is imaginary.

CHAPTER XI.

APPLICATION TO GAMES OF CHANCE AND STATISTICAL PROBLEMS.

76. Correlate between Theory and Practice.—In the theoretical analysis just completed we treated the fundamental elementary functions in the theory of probabilities, the probability function, the expected or probable value of a variable quantity, the mean error, the dispersion and the coefficient of disturbancy. The formulas thus derived were founded upon certain hypothetical axioms, which formed the basis of a mathematical a priori probability as defined by Laplace. As far as the purely abstract mathematical analysis is concerned it matters but little if the hypotheses are physically true or not, that is to say, if they agree with physical facts in the universe as it is known to us. A mathematical analysis may be made on the basis of widely divergent hypotheses, a fact which is clearly shown in the Euclidean and Non-Euclidean geometries. It is, however, quite a different matter when we wish to apply our theory to actual phenomena (physical observed events) as it is evident that a correlation between hypothesis and actual facts follows by no means a priori. It is, of course, true that the different hypotheses in the theory of probabilities are derived to greater or less extent from outside sense data. Such sense data, however, give us only the effect and no clue whatsoever to the relation between cause and effect. In the application of our theory every hypothesis—or rather the results derived from such hypothesis—must be verified by actual experience. Before such a verification is made, we advise the reader to be sceptical and not trust too much in the authority of others but follow the sound advice of Chrystal: "In mathematics let no man over-persuade you. Another man's authority is not your reason." We can so much more encourage an attitude of scepticism in view of the fact that even among the leading mathematicians of the present time there exists no uniform opinion as to the truth of the axioms underlying the theory of probabilities.

77. Homograde and Heterograde Series. Technical Terms. —Whenever a common characteristic or attribute of several groups of observed individual objects or events allows a purely quantitative determination, it may be made the subject of a mathematical analysis and in such cases we are often able to make excellent use of the theory of probabilities. Such quantitative measurements may be divided into various domains of classification. Traces of such classification are found in almost every treatise on mathematical statistics but a uniform system nomenclature is unfortunately lacking among the various statisticians and any one reading the modern literature on mathematical statistics notices often various inconsistencies of the different authors. Mr. G. Udny Yule in his excellent treatise "Theory of Statistics" classifies the statistical series into "statistics of attributes" and "statistics of variables." Apart from the fact that Mr. Yule's statistics of variables also is a statistics of attributes—although of different grades—the author apparently ignores the criterion of Lexis and the associated criterion of Charlier. The German writers use the terms "stetige und unstetige Kollektivgegenstand" (continuous and discontinuous collective objects), which were originally introduced by Fechner. Other writers, such as Johannsen of Denmark and Davenport of America, use still other terms. After having made a comparison of the various systems of classification I have in the following decided to adhere to the system of Charlier wherein the observed statistical series are classified as *homograde* and *heterograde*.

If the individuals all possess the same character or attribute in the same grade (intensity)—or if we disregard the different grades of the attributes—such individuals are called *homograde*, and the statistical series thus formed is a homograde series. If on the other hand we take into consideration the different varying grades of the attributes observed or measured and form the series accordingly we obtain a *heterograde* series. As examples of homograde series we may mention the observed recorded series of coin tossing, card drawings in reference to a specified event, number of births or deaths in a population group, etc. A coin when tossed will either show head or tail, a person will

either be dead or alive. There are no intermediate degrees as for instance that of a half dead person. In all such series the dividing line between the occurrence of the event (attribute) E and the occurrence of the opposite event $\bar{E}$ is distinct and suggests itself a priori and there is no doubt as to the classification of the observed event.

The original record of observation of a homograde series—also known as the *primary list*—is simply a record of the presence or non-presence of a specified attribute of the individuals belonging to the group under observation and is of the following form:

PRIMARY LIST OF HOMOGRADE INDIVIDUALS.

	Attribute.	
Symbol for the Individual	Present (E).	Non-present ($\bar{E}$).
I_1	1	
I_2		1
I_3		1
I_4	1	
I_5	1	
.	.	.
.	.	.
.	.	.

In this scheme the individuals I_1, I_4 and I_5 possess the attribute E while the individuals I_2 and I_3 do not have this attribute.

In observing the presence of a specified attribute in a group of individual objects we meet, however, frequently series of quite another nature than the simple homograde series. When investigating the different measures of heights of persons inside a certain population group no simple dichotomous (*i. e.*, cutting in two) division in two opposite and mutually exclusive groups suggests itself a priori. It is of course true that we might divide the total population under observation into two subsidiary groups of *tall* individuals and *short* individuals. But the question then immediately arises, What constitutes a short or a tall person? The answer must necessarily be arbitrary. Persons above the height of 170 cm. may be classed as tall while persons falling short of such measure may be classed as short persons, and we might in this way form a primary homograde table of the form as given above. There is no logical reason, however, to choose the quantity 170 cm. as the dividing line and comparatively

little value would result from such a classification. It is evident that all persons belonging to groups of talls or shorts are not identical as to the particular attribute in question. The height is merely a characteristic which varies with each individual and no two individuals have mathematically speaking the same height. If we take into consideration the different grades of height among the individuals and arrange the primary table accordingly we obtain a heterograde series of observations. The general form of the primary table of such series is:

PRIMARY LIST OF HETEROGRADE INDIVIDUALS.

Symbol for the Individual.	Grade of Attribute.
I_1	x_1
I_2	x_2
I_3	x_3
I_4	x_4
I_5	x_5
.	.
.	.
.	.
I_n	x_n

Here the quantities $x_1, x_2, \cdots x_n$ give the measures (in kilogram, liter, meter, etc.) of the characteristic in question.[1]

As examples of heterograde series we may mention the lengths, volumes or weights of animals, plants or inorganic objects; astronomical observations as to the brightness of celestial objects; meteorological records of rainfall, temperature or barometer heights; the frequency of deaths among policyholders as to attained age in an assurance company; duration of sickness or disablement, etc.

The investigation of heterograde series is a problem of which we shall treat later under the theory of errors or frequency curves. The homograde series may, however, be explained fully by means of the Bernoullian, Poisson and Lexian Series as founded on the mathematical theory of probabilities in the previous chapters.

78. Computation of the Mean and the Dispersion in Practice. —It would be superfluous to enter into a detailed demonstration of the practical calculation of either the mean or the dispersion

[1] It is to be noted that in the homograde series the primary list is given by abstract numbers while the heterograde series consists of concrete numbers.

were it not for the fact that this calculation is performed with a lot of unnecessary and useless labor by the untrained student and even by many professional statisticians. By the ordinary school method the number zero is chosen as the starting point and all the variables are expressed in their absolute magnitudes, *i. e.*, their distance from 0. In this way one often encounters multiplication and addition of large numbers. The Danish biologist and statistician, W. Johannsen, has illustrated the futility of this method in the following example taken from his treatise "Forelæsninger over Læren om Arvelighed" (Copenhagen, 1905).[1] Dr. Petersen, the director of the Danish Biological Station, counted the tail fin rays of 703 flounders (*Pleuronectes*) caught around the neighborhood of the Skaw. The observations follow:

Number of rays:	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
No. of flounders:	5	2	13	23	58	96	134	127	111	74	37	16	4	2	1

The ordinary way of computing the mean would be as follows:

$$[5 \times 47 + 2 \times 48 + 13 \times 49 + \cdots + 1 \times 61] \div 703,$$

where 703 is the total number of individuals under observation. In Chapter X we gave the following formula for the mean:

$$M = \frac{m_1 + m_2 + m_3 + \cdots + m_N}{N}. \tag{1}$$

This formula may evidently be written as follows:

$$\begin{aligned} M = {} & \frac{m_1 - M_0 + m_2 - M_0 + m_3 - M_0 + \cdots + m_N - M_0}{N} \\ & + M_0 = \frac{\Sigma(m_\nu - M_0)}{N} + M_0 = b + M_0. \end{aligned} \tag{2}$$

In this expression M_0, which Charlier calls the provisional mean, is an arbitrarily chosen number. To show how the introduction of this quantity actually shortens the calculation of the mean we return to the above quoted series of observations of tai: fin rays of flounders.

[1] German edition "Elemente der exakten Erblichkeitslehre" (Jena, 1913), page 11.

NUMBER OF RAYS (x) IN 703 FLOUNDERS ACCORDING TO OBSERVATIONS OF DR. PETERSEN.

$$N = \Sigma F(x) = 703, \quad M_0 = 53.$$

x.	Frequency $= F(x)$	$x - M_0$.	$(x - M_0)F(x)$.	
47	5	-6	-30	
48	2	-5	-10	
49	13	-4	-52	
50	23	-3	-69	
51	58	-2	-116	
52	96	-1	-96	
53	134	$+0$		$+0$
54	127	$+1$		$+127$
55	111	$+2$		$+222$
56	74	$+3$		$+222$
57	37	$+4$		$+148$
58	16	$+5$		$+80$
59	4	$+6$		$+24$
60	2	$+7$		$+14$
61	1	$+8$		$+8$
Sum $= \Sigma$	703		-373	$+845$

We have now:

$$b = (845 - 373) \div 703 = 0.67, \quad M = M_0 + b = 53.67.$$

The method is quite simple and needs hardly any explanation. From a cursory examination of the material we notice that the mean is situated in the neighborhood of the series consisting of 53 rays. We choose therefore the provisional mean, M_0, as 53. We next form the algebraic differences of $x - M_0$. These differences are then multiplied by $F(x)$. The algebraic sum of these products divided by $N = \Sigma F(x)$ gives us the value of b, which quantity added to M_0 gives the value of the mean, M.

To show a slightly modified form of the method we take the following observations of coal-mine accidents in Belgium, covering the period 1901–1910, from "Annales des Mines de Belgique." These data I have reduced to a stationary population group of 140,000 mine workers. In other words the quantity s as defined in § 83 is equal to 140,000.

NUMBER (m) OF PERSONS KILLED IN COAL MINE ACCIDENTS IN BELGIUM, 1901–1910.

$s = 140{,}000, \quad N = 10, \quad M_0 = 140.$

Year.	m.	$m - M_0$.		$(m - M_0)^2$.
1901	164		+24	576
1902	150		+10	100
1903	160		+20	400
1904	130	−10		100
1905	127	−13		169
1906	133	− 7		49
1907	144		+ 4	16
1908	150		+10	100
1909	133	− 7		49
1910	133	− 7		49
Sum = Σ		−44	+68	1608

Hence

$$b = (68 - 44) \div 10 = 2.4, \quad M = 140 + 2.4 = 142.4.$$

In this example probably it would have been easier to have formed the sum Σm_ν directly and then obtained the mean by division by 10. The actual formation of the algebraic sums of $m_\nu - M_0$ however, greatly facilitates the calculation of the dispersion, σ, to which we now shall turn our attention.

The formula for the dispersion

$$\sigma^2 = \frac{\Sigma(m_\nu - M)^2}{N} \quad (\nu = 1, 2, 3, \cdots N) \tag{3}$$

may evidently be written as follows:

$$\begin{aligned}\sigma^2 &= \frac{(m_1 - M_0)^2 + (m_2 - M_0)^2 + \cdots + (m_N - M_0)^2}{N} - b^2 \\ &= \frac{\Sigma(m_\nu - M_0)^2}{N} - b^2,\end{aligned} \tag{4}$$

where b as usual means $M - M_0$, M_0 being the provisional mean. For Belgian coal mine accidents we thus obtain from the above data:

$$\sigma^2 = (1608 \div 10) - 5.76 = 155.04.$$

Where the number of observed individuals is very large an arrangement as that given above for the Belgian statistics becomes too bulky and it is therefore customary to group the observations in classes as for instance in the example of Dr. Johannsen. The dispersion is then computed according to the following elegant

method due to Charlier from whose brochure "Grunddragen af den matematiska Statistiken" ("Rudiments of Mathematical Statistics") I take the following example:

NUMBER OF BOYS (m) PER 500 CHILDREN BORN IN 24 PROVINCES OF SWEDEN DURING EACH MONTH IN 1883 AND 1890.

$s = 500, \quad N = 576, \quad M_0 = 257, \quad w = 5.$

Class.					
Limits m.	Number. $= x$.	Frequency $= F(x)$.	$xF(x)$.	$x^2F(x)$.	$(x+1)^2F(x)$.
200–204	−11	1	− 11	+ 121	100
205–209	−10	0	0	0	0
210–214	− 9	0	0	0	0
215–219	− 8	1	− 8	+ 64	49
220–224	− 7	2	− 14	+ 98	72
225–229	− 6	5	− 30	+ 180	125
231–234	− 5	13	− 65	+ 325	208
235–239	− 4	18	− 72	+ 288	162
240–244	− 3	47	−141	+ 423	188
245–249	− 2	60	−120	+ 240	60
250–254	− 1	81	− 81	+ 81	0
255–259	0	108	0	0	108
260–264	+ 1	91	+ 91	+ 91	364
265–269	+ 2	60	+120	+ 240	540
270–274	+ 3	44	+132	+ 396	704
275–279	+ 4	22	+ 88	+ 352	550
280–284	+ 5	16	+ 80	+ 400	576
285–289	+ 6	6	+ 36	+ 216	294
290–294	+ 7	0	0	0	0
295–299	+ 8	0	0	0	0
300–304	+ 9	1	+ 9	+ 81	100
Sum = Σ		576	+ 14	+3596	+4200

The class width interval in the above scheme was chosen as 5. The observed frequencies are given in column 3. We thus find that the greatest frequency of 108 falls in the class interval 255–259. Choosing this class interval as the origin we designate the other class intervals with their proper positive and negative numbers as shown in column 2. The provisional mean, M_0, is taken as the center of class O, or $M_0 = 257$. In this way the class interval $w = 5$ is taken as the unit.

The whole calculation is very simple. We first of all form the product $x \times F(x)$. The sum of these products divided with $576 = N$ gives the distance—b—from the provisional mean to the arithmetic mean, expressed in units of the class interval, w.

We have thus:

$$b = w \times 14 \div 576 = + 0.0243w = + 0.122,$$

or

$$M = 257 + b = 257.12.$$

The formula for the dispersion takes the form

$$\sigma^2 = w^2 \left[\frac{\Sigma x^2 F(x)}{N} - b^2 \right],$$

where b is expressed in units of the class interval. The table gives us

$\Sigma F(x)x^2 = 3596$ or

$$\sigma^2 = w^2[3596 \div 576 - (0.024)^2] = w^2 6.242,$$

$$\sigma = w \times 2.498 = 12.49.$$

Charlier now checks the results by means of the following relation:

$$\Sigma(x + 1)^2 F(x) = \Sigma x^2 F(x) + 2\Sigma F(x) + \Sigma F(x).$$

For the above example we have:

$$\begin{aligned} \Sigma x^2 F(x) &= + 3{,}596 \\ 2\Sigma x F(x) &= + \quad 28 \\ \Sigma F(x) &= + \quad 576 \\ \hline \text{Sum} &= + 4{,}200 = \Sigma(x + 1)^2 F(x), \end{aligned}$$

which proves the accuracy of the calculation.

The full elegance of the Charlier self checking scheme is shown at a later stage under the calculation of the parameters of frequency curves. In the meantime the student may test the advantage of the provisional mean by trying to compute the mean and the dispersion by the conventional school method. A direct computation by this method would in the last example take about a whole day's labor.

Before we proceed to apply the formulas previously demonstrated, we wish to call the attention of the reader to the following important properties of the mean and the dispersion:

1. The algebraic sum of the deviations from the mean—*i. e.*, $\Sigma(m_\nu - M)$—is zero. This follows immediately from formula (2) of § 78. We have:

$$M = \frac{\Sigma(m_\nu - M_0)}{N} + M_0 = b + M_0,$$

where M_0, the provisional mean, is an arbitrarily chosen number and $b = \Sigma(m_\nu - M_0) \div N$. If $M_0 = M$ we have evidently $b = 0$, which proves the statement.

2. The dispersion (standard deviation) is the least possible root-mean-square deviation, *i. e.*, the root-mean-square deviation is a minimum, when the deviations are measured from the mean.

We have (see formula (4)):

$$\sigma^2 = \frac{\Sigma(m_\nu - M)^2}{N} = \frac{\Sigma(m_\nu - M_0)^2}{N} - b^2,$$

from which the proposition follows *a fortiori*.

79. Westergaard's Experiments.—The Danish statistician, Harald Westergaard, in his "Statistikens Teori i Grundrids" gives the following results of 10,000 observations divided into 100 equal sample sets of drawings of balls from a bag containing an equal number of red and white balls (the ball was returned to the bag after each drawing):

White:	33	34	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Frequency.	0	1	1	2	2	2	3	3	4	5	6	5	11	9	5	10	4	8
White:	55	56	57	58	59	60	61	62	63									
Frequency:	3	5	4	4	0	0	1	1	1.									

The elements as resulting from Westergaard's drawings clearly represent a Bernoullian Series where the number of comparison s is equal to 100. Arranging the data in classes—taking 3 as the class interval—the computation of the mean and the dispersion is easily performed by means of the Charlier self checking scheme.

BERNOULLIAN SERIES. NUMBER OF WHITE BALLS IN 100 DRAWINGS (WESTERGAARD).

$s = 100,\ N = 100,\ M_0 = 49,\ w = 3.$

m.	x.	$F(x)$.	$xF(x)$.	$x^2F(x)$.	$(x+1)^2F(x)$.
33-35	−5	1	− 5	25	16
36-38	−4	0	0	0	0
39-41	−3	5	−15	45	20
42-44	−2	8	−16	32	8
45-47	−1	15	−15	15	0
48-50	0	25	0	0	25
51-53	+1	19	+19	19	76
54-56	+2	16	+32	64	144
57-59	+3	8	+24	72	128
60-62	+4	2	+ 8	32	50
63-65	+5	1	+ 5	25	36
Sum		100	(−51+88)	329	503

CONTROL CHECK.

$$\begin{aligned}\Sigma x^2F(x) &= 329\\ 2\Sigma xF(x) &= 74\\ \Sigma F(x) &= 100\end{aligned}$$

$$\text{Sum} = 503 = \Sigma(x+1)^2F(x)$$

$$b = w(88 - 51) : 100 = w \times 0.37 = 1.11,$$

$$\text{or} \quad M = M_0 + b = 50.11,$$

$$\sigma^2 = w^2[329 : 100 - b^2]^1 = w^2(3.29 - 0.137) = 28.377,$$

$$\text{or} \quad \sigma = 5.33.$$

Giving due allowance for the respective mean errors of the mean and the deviation we have finally:[2]

$$M = 50.11 \pm 0.536, \quad \sigma = 5.33 \pm 0.378.$$

We shall now compare these values with the corresponding theoretical values of the Bernoullian series. The a priori probabilities of drawing red and white are in this example $p = q = \frac{1}{2}$. Hence we have as the theoretical values for the mean and the dispersion:

$$M_B = 100 \times \tfrac{1}{2} = 50, \quad \sigma_B = \sqrt{100 \times \tfrac{1}{2} \times \tfrac{1}{2}} = 5.$$

A comparison between the observed and the theoretical ideal values—taking into account the proper mean errors—shows a very close agreement as far as the dispersion is concerned while the difference in the mean is about $\frac{1}{5}$ of the mean error. A computation of the Lexian Ratio and the Charlier Coefficient of Disturbancy yields the following results:

$$L = 1.072; \quad 100\rho = 3.68.$$

Taking into account the proper mean errors due entirely to the *fluctuation of sampling* we find, however, that our theoretical results and formulas of the previous chapters have been verified in an absolutely satisfactory manner.

80. Charlier's Experiments.—In the above mentioned brochure, "Grunddragen," Charlier gives the results of a long series of card drawings illustrating the Bernoullian, the Poisson and the Lexian Series. As an example showing the frequency dis-

[1] b is expressed in units of w.

[2] For mean errors of M and σ see Addenda.

tribution in a Bernoullian Series Charlier made 10,000 individual drawings (with replacements) from an ordinary whist deck and recorded the number of black and red cards drawn in this manner. Arranging the drawings in sample sets of 10 individual drawings, M. Charlier gives the following table:

BERNOULLIAN SERIES. NUMBER (m) OF BLACK CARDS IN SAMPLE SETS OF 10.

$s = 10,\quad N = 1{,}000,\quad M_0 = 5,\quad w = 1.$

m.	x.	$F(x)$.	$xF(x)$.	$x^2F(x)$.	$(x+1)^2F(x)$.
0	−5	3	− 15	+ 75	+ 48
1	−4	10	− 40	+ 160	+ 90
2	−3	43	−129	+ 387	+ 172
3	−2	116	−232	+ 464	+ 116
4	−1	221	−221	+ 221	0
5	0	247	0	0	+ 247
6	+1	202	+202	+ 202	+ 808
7	+2	115	+230	+ 460	+1,035
8	+3	34	+102	+ 306	+ 544
9	+4	9	+ 36	+ 144	+ 225
10	+5	0	0	0	0
Sum:		1,000	− 67	+2,419	+3,285

CONTROL CHECK.

$$\Sigma^2 xF(x) = +\ 2{,}419$$
$$2\Sigma xF(x) = -\ \ 134$$
$$\Sigma F(x) = +\ 1{,}000$$
$$\text{Sum} = +\ 3{,}285 = \Sigma(x+1)^2F(x)$$

From the above values we obtain:

$b = -\ 67 : 1{,}000 = -\ 0.67;\ \sigma^2 = 2{,}419 : 1{,}000 - b^2 = 2.415.$

Making due allowance for mean errors we have thus:

$$M = 5 - 0.067 = 4.933 \pm 0.050;\quad \sigma = 1.554 \pm 0.035.$$

For the theoretical mean and dispersion we obtain the following values: ($p = q = \frac{1}{2}$)

$$M_B = 5;\quad \sigma_B = 1.581,$$

which gives the following values for the Lexian Ratio and the Charlier coefficient:

$$L = .983,\quad 100\rho \text{ is imaginary.}$$

These results would indicate a slightly subnormal series. Taking into account the fluctuations due to sampling and for which the

mean error serves as a measure the results become normal and serve again as a verification of the theory.

Poisson Series.—As an illustration of the frequency distribution in a Poisson Series Charlier made the following experiment: From an ordinary whist deck was drawn a single card and the color noted. Before the second drawing a spade was eliminated from the deck and replaced by a heart from another deck of cards, so that the deck then contained 12 spades, 13 clubs, 13 diamonds and 14 hearts; from this deck another card was drawn and the color noted. Then another spade was eliminated and a heart substituted. From this deck, containing 11 spades, 13 clubs, 13 diamonds and 15 hearts, a card was again drawn. The drawings were in this manner continued until all the spades were replaced by hearts. The same operation was applied to the clubs, which were replaced by diamonds. After 27 drawings the deck contained only red cards. Altogether 100 sample sets of 27 drawings were made with the following results:

POISSON SERIES. NUMBER (m) OF BLACK CARDS IN SAMPLE SETS OF 27.

$s = 27, \quad N = 100, \quad M_0 = 7, \quad w = 1.$

m.	x.	$F(x)$.	$xF(x)$.	$x^2F(x)$.	$(x+1)^2F(x)$.	Control Check.
3	−4	2	− 8	+ 32	18	
4	−3	6	−18	+ 54	24	+378
5	−2	14	−28	+ 56	14	+ 32
6	−1	14	−14	+ 14	0	+100
7	0	22	0	0	22	
8	+1	17	+17	+ 17	68	+510
9	+2	14	+28	+ 56	126	
10	+3	8	+24	+ 72	128	
11	+4	1	+ 4	+ 16	25	
12	+5	1	+ 5	+ 25	36	
13	+6	1	+ 6	+ 36	49	
	Sum:	100	+16	+378	510	

The calculation of the mean and the dispersion with their respective mean errors yields the following result:

$$b = +0.16, \quad M = 7.16 \pm 0.211,$$
$$\sigma^2 = 3.78 - (0.16)^2 = 3.754, \quad \sigma = 1.937 \pm 0.149.$$

The theoretical Poisson values according to the formulas of § 67 are:

$$M_P = 6.75, \quad \sigma_P = 2.111.$$

If we now take the arithmetic mean of the various proba-

bilities of drawing a black card we find that $p_0 = \frac{1}{4}$. If all the drawings had been performed with a constant probability we should according to the Bernoullian scheme have:

$$M_B = 27 \times \tfrac{1}{4} = 6.75, \quad \sigma_B = \sqrt{27 + \tfrac{1}{4} \times \tfrac{3}{4}} = 2.25.$$

These results verify the formulas as obtained under the discussion of the Poisson Series. ($M_P = M_B$, $\sigma_P < \sigma_B$.)

Lexian Series.—In testing the Lexian Series Charlier first took 10 samples of 10 individual drawings in each sample from an ordinary whist deck. The number of black cards thus drawn was recorded. After this, 10 samples of the same magnitude were taken from a deck containing 25 black and 27 red cards; and then 10 samples from a deck with 24 black and 28 red cards. Of the total 270 samples (until the deck contains only red cards) Charlier gives the first 100 which gave the following result:

LEXIAN SERIES. NUMBER (m) OF BLACK CARDS IN 10 DRAWINGS.
$s = 10$, $N = 100$, $M_0 = 4$.

m.	x.	$F(x)$.	$xF(x)$	$x^2F(x)$.	$(x+1)^2F(x)$.	Control Check.
1	−3	4	−12	+ 36	+ 16	
2	−2	9	−18	+ 36	+ 9	
3	−1	19	−19	+ 19	+ 0	
4	0	21	0	0	+ 21	
5	+1	23	+23	+ 23	+ 92	
6	+2	10	+20	+ 40	+ 90	+294
7	+3	12	+36	+108	+192	+ 76
8	+4	2	+ 8	+ 32	+ 50	+100
	Sum:	100	+38	+294	+470	+470

The final computations (with mean errors) give:

$$b = +\,0.38, \quad M = 4.38 \pm 0.167,$$

$$\sigma^2 = 294 : 100 - b^2 = +\,2.796, \quad \sigma = +\,1.672 \pm 0.118.$$

The mean probability in all trials was:

$$p_0 = 21.50 : 52 = 0.4{,}135, \quad \text{or} \quad M_B = sp_0 = 4.135,$$

$$\sigma_B = \sqrt{sp_0q_0} = 1.557.$$

A calculation of the mean and the dispersion according to the formulas under the Lexian Series (see § 74) gives according to Charlier:

$$M_L = 4.135, \quad \sigma_L = 1.643.$$

This shows that the dispersion in a Lexian Series is greater than the corresponding Bernoullian dispersion. The Lexian Ratio: $L = \sigma_L : \sigma_B$ has the value 1.06. The series according to the terminology of Lexis has a hypernormal dispersion, although a very small one. Charlier in "Grunddragen" (§ 30) says that when arranging the material in 27 samples, each sample containing 100 single trials, the Lexian Ratio has the value $L = 3.82$, indicating a greater hypernormal dispersion than in the smaller samples.

81. Experiments by Bonynge and Fisher.—As an additional verification of the Bernoullian, Poisson and Lexian Series my co-editor, Mr. Bonynge, and myself have repeated the experiments of Westergaard and Charlier in a slightly modified form.

Bernoullian Series.—In 20 sample sets, each set containing 500 individual drawings, from an ordinary whist deck, I counted the number of diamonds drawn in each sample. My records gave the following scheme:

Bernoullian Series. Number of Diamonds (m) in 20 Sample Sets of 500 Drawings.

$s = 500, \quad N = 20, \quad M_0 = 125.$

	$m.$	$m - M_0.$		$(m - M_0)^2.$
	123	-2		4
	143		$+18$	324
	124	-1		1
	133		$+8$	64
	142		$+17$	289
	130		$+5$	25
	117	-8		64
	122	-3		9
	132		$+7$	49
	109	-16		256
	130		$+5$	25
	139		$+14$	196
	138		$+13$	169
	129		$+4$	16
	136		$+11$	121
	121	-4		16
	135		$+10$	100
	124	-1		1
	135		$+10$	100
	116	-9		81
Sum:		-44	$+122$	1,910

The results with their respective mean errors are as follows:

$$M = 128.9 \pm 2.01, \quad \sigma = 8.962 \pm 1.416$$

The theoretical Bernoullian mean and the dispersion have the values:

$$M_B = 125, \quad \sigma_B = \sqrt{spq} = \sqrt{500 \times \tfrac{1}{4} \times \tfrac{3}{4}} = 9.682,$$

where $p = \frac{1}{4}$ denotes the a priori probability of drawing a diamond.

Again I counted the number of aces (irrespective of color) which appeared in 100 sample sets of 100 individual drawings from the same deck of cards. The records arranged in classes gave the following scheme:

NUMBER OF ACES (m) IN 100 SAMPLE SETS OF 100 INDIVIDUAL DRAWINGS.

$s = 100,\quad N = 100,\quad M_0 = 8,\quad w = 1.$

m.	x.	$F(x)$.	$xF(x)$.	$x^2F(x)$.	$(x+1)^2F(x)$.	Control Check
2	−6	1	− 6	36	25	
3	−5	8	−40	200	128	
4	−4	8	−32	128	72	
5	−3	7	−21	53	28	
6	−2	9	−18	36	9	
7	−1	21	−21	21	0	
8	0	13	0	0	13	
9	+1	15	+15	15	60	
10	+2	3	+ 6	12	27	
11	+3	9	+27	81	144	+811
12	+4	1	+ 4	16	25	−110
13	+5	2	+10	50	72	+100
14	+6	2	+12	72	98	
15	+7	0	+ 0	0	0	801
16	+8	0	+ 0	0	0	
17	+9	1	+ 9	81	100	
Sum:		100	−55	811	801	

$$b = -55 : 100 = -0.55,$$

$$M = M_0 + b = 7.45 \pm 0.279 \text{ (with mean error)},$$

$$\sigma^2 = w^2 \frac{\Sigma x^2 F(x)}{\Sigma F(x)} - b^2 = 7.8075,$$

or

$$\sigma = 2.794 \pm 0.198 \text{ (with mean error)}.$$

The theoretical Bernoullian values are:

$$M_B = 100 \times \tfrac{1}{13} = 7.69, \quad \sigma_B = \sqrt{100 \times \tfrac{1}{13} \times \tfrac{12}{13}} = 2.663.$$

A comparison between the empirical and the theoretical a priori values exhibits a close correspondence.

Poisson Series.—As an illustration of the Poisson Series Mr. Bonynge made the following experiment. A sample set of 20 single drawings of balls from an urn (one ball being drawn at a time) was made under the following conditions:

In	drawing	No. 1	the	urn	contained	20	white	and	20	black	balls.
"	"	" 2	"	"	"	21	"	"	19	"	"
"	"	" 3	"	"	"	22	"	"	18	"	"
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
"	"	" 20	"	"	"	39	"	"	1	"	"

Altogether Bonynge took 500 sample sets which arranged in classes give the following scheme:

POISSON SERIES. NUMBER OF BLACK BALLS (m) IN 500 SAMPLE SETS OF 20 INDIVIDUAL DRAWINGS (BONYNGE).

$s = 20, \quad N = 500, \quad M_0 = 5.$

m.	x.	$F(x)$.	$xF(x)$.	$x^2F(x)$.	$(x+1)^2F(x)$.
0	-5	2	$-\ 10$	50	32
1	-4	9	$-\ 36$	144	81
2	-3	35	-105	315	140
3	-2	52	-104	208	52
4	-1	86	$-\ 86$	86	0
5	0	109	0	0	109
6	$+1$	85	$+\ 85$	85	340
7	$+2$	69	$+138$	276	621
8	$+3$	30	$+\ 90$	270	480
9	$+4$	16	$+\ 64$	256	400
10	$+5$	6	$+\ 30$	150	216
11	$+6$	1	$+\ \ 6$	36	49
Sum:	$\Sigma =$	500	$+\ 72$	1876	2520

Hence we have:

$$b = 0.144, \quad M = 5.144, \quad \sigma^2 = 3.732, \quad \sigma = 1.932.$$

The theoretical Poisson values are:

$$M_P = 5.25, \quad \sigma_P = 1.86 \text{ (see formulas, § 74).}$$

The mean of the various probabilities of drawing a black ball is $p_0 = \frac{21}{80}$. According to the Bernoullian scheme we should then have the following values for the mean and the dispersion:

$$M_B = 20 \times \tfrac{21}{80} = 5.25, \quad \sigma_B = (20 \times \tfrac{21}{80} \times \tfrac{59}{80})^{\frac{1}{2}} = 1.968.$$

These values confirm the Poisson theorems ($M_P = M_B, \sigma_P < \sigma_B$).

Lexian Series.—As additional illustration of the Lexian Series

I took 20 sample sets, each set containing 500 drawings of a single ball from an urn (with replacements). The contents of the urn varied from set to set as follows:

Sample set No.			1 : 20	white and		20	black	balls.
"	"	"	2 : 21	"	"	19	"	"
"	"	"	3 : 22	"	"	18	"	"
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
"	"	"	20 : 39	"	"	1	"	"

In the 21st set all the black balls were eliminated and the urn contained white balls only. This set, however, was not taken in consideration in calculating the mean and the dispersion.

LEXIAN SERIES. NUMBER (m) OF BLACK BALLS IN 20 SAMPLE SETS OF 500 INDIVIDUAL DRAWINGS (FISHER).

$s = 500, \quad N = 20, \quad M_0 = 130.$

No. of Set.	m.	$(m - M_0)$.	$(m - M_0)^2$.
1	251	+ 121	14641
2	246	+ 116	13456
3	222	+ 92	8464
4	216	+ 86	7396
5	193	+ 63	3969
6	176	+ 46	2116
7	183	+ 53	2809
8	173	+ 43	1849
9	156	+ 26	676
10	135	+ 5	25
11	140	+ 10	100
12	127	− 3	9
13	115	− 15	225
14	96	− 34	1156
15	78	− 52	2704
16	69	− 61	3721
17	55	− 75	5625
18	43	− 87	7569
19	29	− 101	10201
20	19	− 111	12321
Sum:	Σ =	− 539 + 661	99012

$$b = (661 - 539) : 20 = 6.6, \quad M = M_0 + b = 136.6 \pm 15.86.$$

$$\sigma^2 = 99012 : 20 - b^2 = 4913.4, \quad \sigma = 70.098 \pm 11.09 \text{ (with mean errors).}$$

The theoretical Lexian values are:

$$M_l = 131.25, \quad \sigma_L = 72.676 \text{ (see § 74).}$$

If the series represented a true Bernoullian Series, we should have

$$M_B = 500 \times \tfrac{21}{80} = 131.25, \quad \sigma_B = \sqrt{500 \times \tfrac{21}{80} \times \tfrac{59}{80}} = 9.839.$$

These values confirm the Lexian Theorem ($M_L = M_B$, $\sigma_L > \sigma_B$).

A computation of the Charlier Coefficient of Disturbancy from the observed values gives:

$$100_\rho = 50.80$$

whereas the theoretical value is 55.38, showing a decidedly hypernormal dispersion, a result which was to be expected since the probabilities of drawing black varies from $\frac{1}{2}$ to $\frac{1}{40}$ in the various sets of samples.

All the above experiments show a completely satisfactory verification of the various theorems of the previous chapters and may perhaps serve as a vindication of the followers of Laplace, who like him hold that an *a priori* foundation for probability judgments is indispensable.

CHAPTER XII.

CONTINUATION OF THE APPLICATION OF THE THEORY OF PROBABILITIES TO HOMOGRADE STATISTICAL SERIES.

82. General Remarks.—In this chapter it is our intention to discuss the application of the theory of probabilities to homograde statistical series with special reference to vital statistics. We owe the reader an apology, however, inasmuch as in the former paragraphs we have employed the term *statistics* without defining its meaning in a rigorous manner. A definition may perhaps appear superfluous since statistics nowadays is almost a household word. The term unfortunately is often employed as a mere phrase without any understanding of its real meaning. This applies especially to that band of self-styled statisticians, mere dilettanti, who, with an energy which undoubtedly could be better employed otherwise, attempt to investigate and analyze mass phenomena regardless of method and system. When investigations are undertaken by such dilettanti the common gibe that "statistics will prove anything" becomes, alas, only too true and proves at least that "like other mathematical tools they can be wielded effectively only by those who have taken the trouble to understand the way they work."[1]

By the science of statistics we understand the recording and subsequent quantitative analysis of observed mass phenomena.

By mathematical statistics (also called statistical methods) we understand the quantitative determination and measurement of the effect of a complex of causes acting on the object under investigation as furnished by previously recorded observations as to certain attributes among a collective body of individual objects.

Practical statistics—if such a name may be used—then simply becomes the mechanical collection of statistical data, *i. e.*, the recording of the observed attributes of each individual. In no way do we wish to underestimate the importance of this process

[1] See Nunn, "Exercises in Algebra" (London, 1914), pages 432–33.

which is as important for the statistical analysis as is the gathering of structural materials for the erection of a large building.

Mathematical statistics is thus the tool we must use in the final analysis of the statistical data. It is a very effective and powerful tool when used properly by the investigator. At the same time it is not an automatic calculating machine in which we need only put the material and read off the result on a dial. A person without any knowledge whatsoever about the nature of logarithms may in a few hours be taught how to use a logarithmic table in practical computations, but it would be foolish to view the formulas and criteria from probabilities when applied to statistical data in the same light as a table of logarithms in calculating work. Such formulas and criteria must be used with caution and discretion and only by those who have taken the trouble to make a thorough study of probabilities and master their real meaning and their relation to mass phenomena. If put in the hands of mere amateurs the formulas become as dangerous a toy as a razor to a child.

It is not our intention to give in this work a description of the technique of the collection of the material, which depends to a large extent on local social conditions and for which it is difficult to give a set of fixed rules. In the following we shall treat the mathematical methods of statistics exclusively, and furthermore make the theory of probabilities the basis of our investigations.

83. Analogy between Statistical Data and Mathematical Probabilities.—Let us for the moment imagine a closed community with a stationary population from year to year and let us denote the size of such a population by s. Let us furthermore suppose we were given a series of numbers:

$$m_1, m_2, m_3, \cdots m_N,$$

denoting the number of children born in various years in this community. The ratios

$$\frac{m_1}{s}, \frac{m_2}{s}, \frac{m_3}{s}, \cdots \frac{m_N}{s}$$

may then be looked upon as probabilities of a childbirth in various years. As Charlier justly remarks, "such an identification of a statistical ratio with a mathematical probability is

at first sight a mere analogy which possibly may have very little in common with the observed statistical phenomena, but a closer scrutiny shows the great importance for statistics of such a view." If such ratios could be regarded as mathematical probabilities wherein the various m's were identical to favorable cases in s total trials, the mean and the dispersion could be determined a priori from the Bernoullian Theorem. The founders of mathematical statistics regarded the identification of an ordinary statistical series with a Bernoullian Series almost as axiomatic. This view is found even among some leading writers of the present time. Among others we apparently find this traditional view by the eminent English actuary, G. King, in his classic "Text Book." In Chapter II of this well-known standard actuarial treatise a probability is defined as follows: "If an event may happen in α ways and fail in β ways, all these ways being equally likely, the probability of the happening of the event is $\alpha \div (\alpha + \beta)$." With this definition as a basis King then deduces the elementary formulas of the addition and multiplication theorems. He then continues: "Passing now to the mortality table, if there be l_x persons living at age x, and if these l_{x+n} survive to age $x + n$, then the probability that a life aged x will survive n years is $l_{x+n} \div l_x = {}_np_x$. And again "the probability that a life aged x and a life aged y will both survive n years is ${}_np_x \times {}_np_y$."[1] From the above it would appear that the author unreservedly assumes a one-to-one correspondence between the l_{x+n} survivors and "favorable ways" as known from ordinary games of chance and a similar correspondence between the original l_x persons and "equally possible cases." A simple consideration will show that there exists no a priori reason for such a unique correspondence between ordinary empirical death rates and mathematical probabilities. None of the original l_x persons can be considered as

[1] Mr. H. Moir in his "Primer of Insurance" tried to avoid the difficulty by giving a wholly new definition of "equally likely events." According to Moir "events may be said to be 'equally likely' when they recur with regularity *in the long run*." Apart from the half metaphorical term "in the long run" Mr. Moir fails to state what he means by the expression "with regularity." If the statement is to be understood as regular repetitions of a certain event in various sample sets, it is evident that we may obtain a regular recurrence of the observed absolute frequencies in a Poisson Series, where—as we know—the events are *not* equally likely."—A.F.

being "equally likely" as in the sense of games of chance. Numerous factors such as heredity, environment, climatic and economic conditions, etc., play here a vital part in the various complexes embracing the original l_x persons.

The belief in an absolute identity of mathematical probabilities and statistical frequency ratios seems to have originated from Gauss. The great German mathematician—or rather the dogmatic faith in his authority as a mathematician—proved thus for a number of years a veritable stumbling block to a fruitful development of mathematical statistics. Gauss and his followers maintained that all statistical mass phenomena could be made to conform with the law of errors as exhibited by the so-called Gaussian Normal Error Curve. If certain statistical series exhibited discrepancies they claimed that such deviations arose from the limited number of observations. The deviations would become less marked if the number of observed values was enlarged and would eventually disappear as the number of observations approached infinity as its ultimate value. The Gaussian dogma held sway despite the fact that the Danish actuary, Oppermann, and the French mathematicians, Binemaye and Cournot, have pointed out that several statistical series, despite all efforts to the contrary offered a persistent defiance to the Gaussian law. The first real attack on the dogma laid down so authoritatively by Gauss was delivered by the French actuary, Dormay, in certain investigations relating to the French census. It was, however, first after the appearance of the already mentioned brochure by Lexis, "Die Massenerscheinungen, etc.," that a correct idea was gained about the real nature of statistical series.

The Lexian theory was expounded in the previous chapters of this work, and we are therefore ready to enter upon the investigations of a few selected mass observations from the domain of vital statistics.

84. Number of Comparison and Proportional Factors.—In the mathematical treatment of the Lexian theory of dispersion we tacitly assumed that the total number of individual trials in a sample set or the number of comparison, s, remained constant from set to set. In the observations on games of chance it

remained in our power to arrange the actual experiments in such a manner that s would be constant. In actual social statistical series such simple conditions do not exist. In comparing the number of births in a country with the total population it is readily noticed that the population does not remain constant but varies from year to year. For this reason the various numbers m denoting the births are not directly comparable with another. We may, however, easily form a new series of the form:

$$\frac{s}{s_1} \cdot m_1, \ \frac{s}{s_2} \cdot m_2, \ \frac{s}{s_3} \cdot m_3, \ \cdots \ \frac{s}{s_N} \cdot m_N,$$

wherein the various numbers, m_1, m_2, m_3 $\cdots$, corresponding to the numbers of comparison s_1, s_2, s_3, $\cdots$, are reduced to a constant number of comparison s. This series is by Charlier called a *reduced statistical series*. Such a reduction requires, in many

PROPORTIONAL FACTORS FOR A HYPOTHETICAL STATIONARY POPULATION IN SWEDEN AND DENMARK EQUAL TO 5,000,000 AND 2,500,000 RESPECTIVELY.

Year.	Sweden, Inhabitants.	$s:s_k$.	Year.	Denmark, Inhabitants.	$s:s_k$
1876	4,429,713	1.1288	1888	2,143,000	1.1666
1877	4,484,542	1.1150	89	2,161,000	1.1569
1878	4,531,863	1.1033	1890	2,179,000	1.1473
79	4,578,901	1.0919	91	2,195,000	1.1390
1880	4,565,668	1.0952	92	2,210,000	1.1312
81	4,572,245	1.0936	93	2,226,000	1.1230
82	4,579,115	1.0919	94	2,248,000	1.1121
83	4,603,595	1.0861	1895	2,276,000	1.0984
84	4,644,448	1.0765	96	2,306,000	1.0841
1885	4,682,769	1.0677	97	2,338,000	1.0694
86	4,717,189	1.0600	98	2,371,000	1.0544
87	4,734,901	1.0560	99	2,403,000	1.0404
88	4,748,257	1.0530	1900	2,432,000	1.0280
89	4,774,409	1.0472	01	2,462,000	1.0154
1890	4,784,981	1.0449	02	2,491,000	1.0036
91	4,802,751	1.0410	03	2,519,000	0.9925
92	4,806,865	1.0402	04	2,546,000	0.9819
93	4,824,150	1.0365	1905	2,574,000	0.9713
94	4,873,183	1.0261	06	2,603,000	0.9604
1895	4,919,260	1.0165	07	2,635,000	0.9488
96	4,962,568	1.0076	08	2,668,000	0.9370
97	5,009,632	0.9981	09	2,702,000	0.9252
98	5,062,918	0.9875	1910	2,737,000	0.9134
1899	5,097,402	0.9809	11	2,800,000	0.8929
1900	5,136,441	0.9734	1912	2,830,000	0.8834

cases, a certain correction. However, when the general ratios $s \div s_k$ ($k = 1, 2, 3 \cdots N$) are close to unity the reduced series may be treated as a directly observed series. In most of the following examples taken from Scandinavian statistical tabular works the *proportional factor* $s \div s_k$, is close to unity as shown in the table below. For Sweden I have, following Charlier, assumed a stationary population $s = 5{,}000{,}000$. The corresponding Danish s I have taken as 2,500,000.

The above figures are taken from "Sveriges officielle statistik" and "Statistisk Aarbog for Danmark" for 1913 (Précis de Statistique, 1913).

85. Child Births in Sweden.—From Charlier's "Grunddragen" I select the following example showing the number of children born in Sweden in the period from 1881–1900 as reduced to a stationary population of 5,000,000.

Number of Children born in Sweden as to Calendar Year (Charlier).
$s = 5{,}000{,}000$, $N = 20$, $M_0 = 140{,}000$.

Year.	m.	$m - M_0$.		$(m - M_0)^2$.
1881	145,230	+5,230		27,352,900
82	146,630	+6,630		44,089,600
83	144,320	+4,320		18,662,400
84	149,360	+9,360		87,609,600
1885	146,600	+6,600		43,560,000
86	148,270	+8,270		68,392,900
87	148,020	+8,020		64,320,400
88	143,680	+3,680		13,542,400
89	138,300		−1,700	2,890,000
1890	139,600		− 400	160,000
91	141,070	+1,070		1,144,900
92	134,830		−5,170	26,728,900
93	136,540		−3,460	11,971,600
94	134,840		−5,160	26,625,600
1895	136,820		−3,180	10,112,400
96	135,330		−4,670	21,808,900
97	132,750		−7,250	52,562,500
98	134,820		−5,180	26,832,400
99	131,320		−8,680	75,342,400
1900	134,460		−5,540	30,691,600
	Sum Σ =	+ 53,190	− 50,390	654,401,400

From which we obtain:

$b = (+\ 53{,}190 - 50{,}390) : 20 = 140$

$M = M_0 + b = 140{,}140$

$$\sigma^2 = 654{,}401{,}400 : 20 - b^2 = 32{,}700{,}470, \text{ or } \sigma = 5{,}718.$$

The empirical probability of a birth (p_0) is

$p_0 = M : s = 0.02803$, so that $q_0 = 1 - p_0 = 0.97197$ and the Bernoullian dispersion

$$\sigma_B = \sqrt{s p_0 q_0} = 369.0.$$

The actual observed dispersion (5,718) is thus much greater than the Bernoullian. The birth series is considerably hypernormal. The Lexian ratio has the value

$$L = 5{,}718 : 369.0 = 15.50,$$

while the Charlier coefficient of disturbancy is:

$$100\rho = 4.07.$$

Both the values of L and ρ show that the birth series by no means can be compared with the ordinary games of chance but is subject to outward perturbing influences.

86. Child Births in Denmark.—The following example shows the corresponding birth series for Denmark in the 25-year period from 1888–1912 as reduced to a stationary population of 2,500,000. The computation of the various parameters follows:

$$b = (39{,}713 - 30{,}287) : 25 = +377,$$

$$M = M_0 + b = 73{,}377,$$

$$\sigma^2 = 281{,}208{,}156 : 25 - b^2 = 11{,}106{,}197.2,$$

$$\sigma_B^2 = s p_0 q_0 = 71{,}223. \quad (p_0 = M : s = 0.0293508),$$

$$L = \sigma : \sigma_B = 12.5$$

$$100\rho = 100(\sqrt{\sigma^2 - \sigma_B^2}) : M = 4.52.$$

NUMBER OF CHILDREN BORN IN DENMARK AS TO CALENDAR YEAR.

$s = 2{,}500{,}000$, $N = 25$, $M_0 = 73{,}000$.

Year.	m.	$m - M_0$.	$(m - M_0)^2$.
1888	78,659	+ 5,659	32,024, 281
89	77,956	+ 4,956	24,561,936
1890	76,154	+ 3,154	9,947,716
91	77,377	+ 4,377	19,158,129
92	74,059	+ 1,059	1,121,481
93	76,965	+ 3,965	15,721,225
94	75,956	+ 2,956	8,740,636
1895	75,649	+ 2,649	7,017,201
96	76,183	+ 3,183	10,131,489
97	74,404	+ 1,404	1,971,216

Year.	m.		m−M₀.	(m−M₀)².
98	75,570		+ 2,570	6,604,900
99	74,236		+ 1,236	1,527,606
1900	74,146		+ 1,146	1,313,316
01	74,341		+ 1,341	1,798,281
02	73,058		+ 58	3,364
03	71,802	− 1,198		1,435,204
04	72,359	− 641		410,881
1905	70,981	− 2,019		4,076,361
06	71,280	− 1,720		2,958,400
07	70,516	− 2,484		6,170,256
08	71,438	− 1,567		2,455,489
09	79,597	− 2,403		5,774,409
1910	68,777	− 4,223		17,833,729
11	66,016	− 6,984		48,776,256
1912	65,952	− 7,048		49,674,304
	Sum: Σ =	−30,287	+39,713	281,208,156

Practically the same deductions hold true for this Danish series as for the Swedish series. We meet again a hypernormal series subject to perturbing influences. The closeness of the two values of the Charlier coefficient of disturbancy indicates that the number of births in Sweden and Denmark apparently are subject to the same outward disturbing influences.

87. Danish Marriage Series.—The following table shows the number of marriages in Denmark from 1888–1912.

Number of Marriages in Denmark.

$s = 2{,}500{,}000,\ N = 25,\ M_0 = 18{,}000.$

Year.	m.	m − M₀.		(m − M₀)².
1888	17,605	− 395		156,025
89	17,622	− 378		142,884
1890	17,181	− 819		670,761
91	17,017	− 983		966,289
92	17,012	− 988		976,144
93	17,676	− 324		104,976
94	17,445	− 555		308,025
1895	17,736	− 264		69,696
96	18,239		+ 239	57,121
97	18,676		+ 676	456,976
98	18,870		+ 870	756,900
99	18,661		+ 661	436,921
1900	19,015		+1,015	1,030,225
01	17,870	− 130		16,900
02	17,712	− 288		82,944
03	17,791	− 209		43,681
04	17,895	− 105		11,025
1905	17,947	− 53		2,809

Year.	m.	$m-M_0$.		$(m-M_0)^2$.
06	18,592		+ 592	350,464
07	19,072		+1,072	1,149,184
08	18,750		+ 750	562,500
09	18,453		+ 453	205,209
1910	18,255		+ 255	65,025
11	17,749	− 251		63,001
1912	18,034		+ 34	1,156
	Sum: Σ =	−5,742	+6,617	8,686,841

Hence we have:

$$b = (6,617 - 5,742) : 25 = 35, \ M = M_0 + b = 18,035.$$

$$\sigma^2 = (8,686,841 : 25) - b^2 = 346,249, \ \sigma = 588.43,$$

$$\sigma_B = 133.81, \ L = 4.41, \ 100\rho = 5.73.$$

We encounter again a hypernormal series with quite large perturbations. For Sweden Charlier has computed the coefficient of disturbancy for marriages in the period 1876–1900 and found it to be 5.49. A comparison with the same quantity for the above Danish data shows that the perturbing influences for the two countries are about the same.

88. Stillbirths.—As another example from vital statistics I give the number of stillbirths in Denmark from 1888–1912 as compared with a hypothetical number of 70,000 births per annum.

NUMBER OF STILLBIRTHS IN DENMARK AS REDUCED TO A STATIONARY NUMBER OF 70,000 BIRTHS PER ANNUM.

$s = 70,000, \ N = 25, \ M_0 = 1,700.$

Year.	m.	$m-M_0$.		$(m-M_0)^2$.
1888	1,861		+ 161	25,921
89	1,924		+ 224	50,176
1890	1,830		+ 130	16,900
91	1,779		+ 79	6,241
92	1,811		+ 111	12,321
93	1,788		+ 88	7,744
94	1,719		+ 19	361
1895	1,753		+ 53	2,809
96	1,714		+ 14	196
97	1,811		+ 111	12,321
98	1,797		+ 97	9,409
99	1,737		+ 37	1,369
1900	1,696	− 4		16
01	1,732		+ 32	1,024
02	1,694	− 6		36
03	1,685	− 15		225
04	1,682	− 18		324

Year.	m.	$M-m_0$.		$(m-M_0)^2$.
1905	1,705		+ 5	25
06	1,620	− 80		6,400
07	1,723		+ 23	529
08	1,694	− 6		36
09	1,665	− 35		1,225
1910	1,658	− 42		1,764
11	1,659	− 42		1,764
12	1,638	− 62		3,844
	Sum: Σ =	−310	+1,184	161,216

Actual computation gives:

$$b = (1{,}184 - 310) : 25 = 34.96,\ M = 1{,}734.96,$$

$$\sigma^2 = 161{,}216 : 25 - b^2 = 5{,}226.44,\ 100\rho = 3.407.$$

The series is again hypernormal. We shall show presently, when discussing the disturbing influences, that this series after the elimination of the secular perturbations actually represents a normal series. In the meantime we give a few examples relating to accident statistics.

89. Coal Mine Fatalities.—The following table gives the number of deaths from accidents in coal mines in various countries in the period 1901–1910 together with the number of comparison s.

Year	Belgium $s = 140{,}000$	Austria $s = 68{,}000$	England $s = 900{,}000$	France $s = 180{,}000$	Germany $s = 500{,}000$	Japan $s = 110{,}000$	United States $s = 610{,}000$
1901	164	81	1,224	218	1,170	263	1,982
02	150	73	1,116	196	995	188	2,263
03	160	50	1,134	184	960	278	1,952
04	130	62	1,116	193	900	239	2,135
1905	127	99	1,215	187	930	354	2,214
06	133	70	1,161	1,262	985	578	2,944
07	144	73	1,179	198	1,240	399	2,977
08	150	58	1,188	171	1,355	262	2,220
09	133	73	1,287	210	1,021	667	2,440
1910	133	63	1,530	194	985	245	2,391

This gives the following values for the Charlier coefficient:

	100ρ
Belgium	2.55
Austria	13.85
England	4.71
France	34.19
Germany	9.27
Japan	44.12[1]
U. S. A.	12.07

[1] I doubt whether the Japanese data as given by the Bureau of Mines are reliable.

The comparatively large values of ρ show that the fatal accidents in coal mines are subject to violent perturbations. The disturbing influences are greatest for France where the Charlier coefficient is above 34, which immediately shows that some powerful disturbing influence has made itself felt. Looking over the table we find a very large number of deaths for the year 1906. The extremely heavy death rate in this year was caused by the Courrieres mine explosion, in which 1,099 persons lost their lives and marks probably the most fatal disaster in the whole history of coal-mining. Eliminating this catastrophe from the data in the table given above we find indeed that the coefficient of disturbancy becomes imaginary, indicating very stable conditions in French mines. Thus eliminating the more fatal catastrophes we get at least for France a subnormal series for the everyday accidents. In order better to illustrate the influence of the elimination of the most disturbing catastrophes I submit the following two series as reduced to a stationary $s = 630,000$ of fatal coal mine accidents in the United States in the period 1900–1914 as recorded by the Bureau of Mines. The first series shows total number of deaths m_k, the second series gives the total deaths m_k' per year after eliminating all such accidents in which 5 or more men were killed.

NUMBER OF DEATHS FROM ACCIDENTS IN COAL MINES IN UNITED STATES.
$s = 630,000$, $N = 15$.

	m_k	m_k'		m_k	m_k'
1900	2,173	1,843	1908	2,293	1,967
01	2,048	1,863	09	2,520	2,053
02	2,337	1,837	1910	2,470	2,085
03	2,016	1,768	11	2,350	1,984
04	2,205	1,911	12	2,060	1,839
1905	2,286	1,964	13	2,350	1,957
06	2,111	2,075	1914	2,070	1,810
07	3,074	2,190			

The first series gives a coefficient of disturbancy equal to 11.06 while the same quantity for the second series has the value 5.51. Despite the fact that the coefficient of disturbancy is reduced about 50 per cent. there still remains disturbing influences, which clearly shows that conditions in American mines are not so stable as in the mines of France, Belgium and England.

90. Reduced and Weighted Series in Statistics.—So far all our problems in statistical analysis have been related to series where the value of s was constant or where the ratio $s : s_k$ was so close to unity that it might be used as a factor of proportionality. We shall now consider the case where this ratio differs greatly from unity. As an illustration of this kind of series I choose the number of fatal coal mine accidents in various states of the American Federation together with the number of people engaged in coal mining in these states. The figures as taken from the report of the Bureau of Mines relate to the year of 1914.[1]

NUMBER OF PERSONS ENGAGED IN MINING (s_k) AND NUMBER KILLED (m_k) IN 20 STATES DURING THE YEAR 1914.

$s = 1000, \quad N = 20.$

	s_k.	m_k.	$p_0 s_k$.	$\lvert m_k - p_0 s_k \rvert$.	
1 Alabama..........	24,552	128	73	55	
2 Colorado..........	10,550	75	31	44	
3 Illinois............	79,529	141	237	96	
4 Indiana...........	22,110	44	66	22	
5 Iowa..............	15,757	37	47	10	
6 Kansas............	12,500	33	37	4	
7 Kentucky.........	26,332	61	79	18	
8 Maryland.........	5,675	18	17	1	
9 Missouri..........	10,418	19	31	12	
10 New Mexico.......	4,021	18	12	6	
11 Ohio..............	45,815	62	136	74	
12 Oklahoma.........	8,948	31	27	24	
13 Pennsylvania......	175,745	595	524	71	(Anthracite Mines)
14 Pennsylvania......	172,196	402	513	111	(Bituminous Mines)
15 Tennessee.........	9,580	26	29	3	
16 Texas.............	4,900	11	15	4	
17 Virginia...........	9,162	27	27	0	
18 Washington.......	5,730	17	17	0	
19 W. Virginia........	74,786	371	223	148	
20 Wyoming.........	8,353	51	25	26	
Sum: Σ =	726,659	2,167		709	

It will be noted that the population engaged in mining varies greatly from state to state. In making a simple reduction to a common number of comparisons by a proportional factor it is evident, however, that we would give the same weight to the observed from New Mexico with a population of miners equal to

[1] Catastrophes in the Eccles Mine in West Virginia and in the Royalton Mine of Illinois are eliminated.

4,021 as to the mining population of the state of Pennsylvania where over 340,000 persons are engaged in the same industry. This procedure is faulty. Let us imagine for the moment two sets of drawings from a bag containing white and black balls. The first sample set contained 10,000 drawings and the second set only 100 drawings. If these series were reduced to a common number of comparison $s = 1{,}000$ we should have

$\frac{1{,}000}{10{,}000} m_1$ and $\frac{1{,}000}{100} m_2$ (m_1 and m_2 standing for the number of white balls) as the number of white balls drawn in sample sets of 1,000 single drawings.

But these values are not equally reliable. The mean error in the second series is in fact 10 times as large as the mean error in the first series. In order to overcome this difficulty we ask the reader to consider the following series:

$$
\begin{array}{lllllll}
\text{The element} & & \frac{s}{s_1} m_1 & \text{is repeated} & & s_1 & \text{times} \\
\text{"} & \text{"} & \frac{s}{s_2} m_2 & \text{"} & \text{"} & s_2 & \text{"} \\
\text{"} & \text{"} & \frac{s}{s_3} m_3 & \text{"} & \text{"} & s_3 & \text{"} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\text{"} & \text{"} & \frac{s}{s_N} m_N & \text{"} & \text{"} & s_N & \text{"}
\end{array}
$$

In this way we obtain a series with $s_1 + s_2 + s_3 + \cdots + s_N$ elements which may be termed a *reduced and weighted series* since the larger s_k appears oftener than the smaller values of s_k. We shall now see if it is possible to determine the expected value of the mean and the dispersion if the series is supposed to follow the Bernoullian Law.

The mean is defined by the following relation:

$$M = \Bigg[\overbrace{\frac{s}{s_1} m_1 + \cdots + \frac{s}{s_1} m_1}^{s_1} + \overbrace{\frac{s}{s_2} m_2 + \cdots + \frac{s}{s_2} m_2}^{s_2}$$

$$+ \cdots + \overbrace{\frac{s}{s_N} m_N + \cdots + \frac{s}{s_N} m_N}^{s_N} \Bigg] \div [s_1 + s_2 + \cdots + s_N]$$

$$= \frac{\Sigma s_k \frac{s}{s_k} m_k}{\Sigma s_k} = \frac{s \Sigma m_k}{\Sigma s_k}.$$

Denoting the average empirical probability by p_0 we have $\Sigma m_k : \Sigma s_k = p_0$ and,

$$M_B = s p_0.$$

As to the dispersion it takes on the following form:

$$\sigma^2 = \Bigg[\overbrace{\left(\frac{s}{s_1} m_1 - s p_0\right)^2 + \cdots + \left(\frac{s}{s_1} m_1 - s p_0\right)^2}^{s_1}$$

$$+ \overbrace{\left(\frac{s}{s_2} m_2 - s p_0\right)^2 + \cdots + \left(\frac{s}{s_2} m_2 - s p_0\right)^2}^{s_2} + \cdots$$

$$+ \overbrace{\left(\frac{s}{s_N} m_N - s p_0\right)^2 + \cdots + \left(\frac{s}{s_N} m_N - s p_0\right)^2}^{s_N} \Bigg]$$

$$\div [s_1 + s_2 + \cdots s_N]$$

$$= \frac{\left(\Sigma s_k \frac{s}{s_k} m_k - s p_0\right)^2}{\Sigma s_k} = \frac{\Sigma \frac{s^2}{s_k} (m_k - s_k p_0)^2}{\Sigma s_k} \quad (k = 1, 2, 3, \cdots N).$$

In finding the theoretical dispersion, assuming a Bernoullian distribution for which p_0 may be used an an approximation of the mathematical a priori probability, we ask the reader to examine the general term of the expression for σ^2, viz.:

$$\left[\frac{s^2}{s_k} (m_k - s_k p_0)^2 \right] : \Sigma s_k.$$

If the individual trials follow the Bernoullian Law the expected value of the factor $(m_k - s_k p_0)^2$ takes the form:

$$e[(m_k - s_k p_0)^2] = \Sigma (m_k - s_k p_0)^2 \varphi(m_k) = s_k p_0 q_0.$$

This brings the general term for σ^2 to the form:

$$\frac{s^2}{\Sigma s_k} p_0 q_0 = \frac{s}{\Sigma s_k} s p_0 q_0.$$

Thus the expected value of σ^2 according to the Bernoullian distribution may be written as follows:

$$\sigma_B{}^2 = \sum_{k=1}^{k=N} \frac{s}{\Sigma s_k} s p_0 q_0 = \frac{Ns}{\Sigma s_k} s p_0 q_0, \quad \text{or}: \sigma_B = f\sqrt{s p_0 q_0},$$

where as before $p_0 = \Sigma m_k : \Sigma s_k$ and $f = \sqrt{\dfrac{Ns}{\Sigma s_k}}$.

These formulas give us the means of computing the Lexian Ratio and the Charlier coefficient of disturbancy in the ordinary way. Some of the computations require, however, a great amount of arithmetical work and the goal is reached more easily by making use of the mean deviation (in § 74*a*). We found there the following relation:

$$\sigma = 1.2533\vartheta.$$

In the weighted series it is readily seen that the value of ϑ will be of the form:

$$\frac{\Sigma s_k \left| \dfrac{s}{s_k} m_k - s p_0 \right|}{\Sigma s_k} = \frac{\Sigma s \left| m_k - s_k p_0 \right|}{\Sigma s_k}.$$

If the series may be assumed to follow a Bernoullian distribution we have

$$\sigma_B = 1.2533\vartheta.$$

From the above formulas it is readily noticed that we may find the mean and the dispersion directly from the observed series without a preliminary reduction to a common number of comparison s. This is in fact the method used in the above example of coal-mine accidents in various states. We have:

$$p_0 = \Sigma m_k : \Sigma s_k = 2{,}167 : 726{,}659 = .002982,$$

$$\vartheta = \frac{\Sigma s \left| m_k - s_k p_0 \right|}{\Sigma s_k} = 1{,}000 \times 709 : 726{,}659 = 0.9757,$$

$$\sigma = 1.2533 \times \vartheta = 1.223,$$

$$\sigma_B{}^2 = f^2 s p_0 q_0 = \frac{20{,}000}{726{,}659} \times 1{,}000 \times 0.997 \times 0.003 = 0.0817,$$

$$100\rho = \frac{100\sqrt{\sigma^2 - \sigma_B{}^2}}{M} = 40 \text{ approx.}$$

The large value of the Charlier coefficient of disturbancy clearly shows that conditions in coal mines by no means are uniform in the whole union but vary greatly according to the locality. An actual computation shows in fact that in a few states such as Michigan and Iowa we find an imaginary coefficient of disturbancy whereas States as Ohio and West Virginia exhibit marked hypernormal series with a large coefficient of disturbancy. The establishment of this fact is of some importance in connection with accident assurance. Many statisticians seem to be of the opinion that a standard accident table computed from the data of the whole union ought to serve as the basis for assurance premiums. Such a table would assume uniform conditions all over the union. The enormously high value of ρ as computed above shows the fallacy of such a view.

91. Secular and Periodical Fluctuations.—In the last paragraphs we have just learned how to detect the presence of disturbing influences in a statistical series. A value of the Lexian ratio differing from unity or a value of the Charlier coefficient of disturbancy differing from zero indicates the presence of fluctuations in the chances for the event or phenomena under investigation. After having established the presence of such fluctuations it is the duty of the statistician to trace the sources of the disturbing influences. This is in general done by means of the theory of correlation, which will be discussed in the second volume of this work.

It is, however, possible to classify the disturbances under two categories which by Charlier are termed as *secular* and *periodical* variations.[1] The periodical fluctuations are in general difficult to discuss on account of the variations in the period of the disturbing forces. In many cases we are in absolute ignorance about the length of such a period and therefore unable to subject the series to a mathematical analysis. If the length of the period is known it is indeed not difficult to determine the periodical disturbances. This is often the case in series giving the occurrence of a certain disease in various months. In statistics giving the frequency of malaria in a community the observed cases are

[1] Lexis uses the terms "evolutionary" ("symptomatic") and "periodical" ("oscilating") for such fluctuations.

nearly all limited to the warmer months and infrequent in the winter months.

In the secular fluctuations due to certain outward influences working continually in the same direction it is quite easy to calculate the rate of such variations.

Let β denote the increase (decrease) of the original probabilities ($p_1, p_2, p_3, \cdots p_n$) from set to set in the given statistical series so that

$$p_2 - p_1 = \beta$$
$$p_3 - p_2 = \beta$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$p_N - p_{N-1} = \beta$$

We then have:

$$p_k = p_1 + (k - 1)\beta. \tag{1}$$

The mean probability has the value:

$$\begin{aligned} p_0 &= \frac{p_1 + p_2 + p_3 + \cdots + p_N}{N} \\ &= \frac{p_1 + p_1 + \beta + p_1 + 2\beta + \cdots + p_1 + (N-1)\beta}{N} \\ &= p_1 + \frac{N-1}{2}\beta. \end{aligned} \tag{2}$$

Eliminating p_1 from (1) and (2) we have:

$$p_k - p_0 = \left(k - \frac{N+1}{2}\right)\beta.$$

If the observed and reduced numbers $m_1, m_2, m_3, \cdots m_N$ may be regarded as approximately coinciding with $sp_1, sp_2, sp_3, \cdots sp_N$ we may write (2) as follows:

$$m_k - M = \left(k - \frac{N+1}{2}\right)s\beta \quad (k = 1, 2, 3, \cdots N). \tag{3}$$

In order to obtain an expression for $s\beta$ in known quantities we must elimate the quantity k. Multiplying both sides of the equation (3) by $k - (N + 1)/2$ we have:

$$(m_k - M)\left(k - \frac{N+1}{2}\right) = \left(k - \frac{N+1}{2}\right)^2 s\beta.$$

Summing this expression for all values k from $k = 1$ to $k = N$ we have:

$$\Sigma\left(k - \frac{N+1}{2}\right)(m_k - M) = s\beta\,\Sigma\left(k - \frac{N+1}{2}\right)^2. \qquad (4)$$

The following expressions from the summation of series are well known to the reader from elementary algebra:

$$\sum_1^N k^2 = \frac{1}{6}N(N+1)(2N+1),$$

$$\sum_1^N k = \frac{1}{2}N(N+1).$$

Substituting these values in (4) we obtain after a few simple transformations the following expression for $s\beta$:

$$s\beta = \frac{12}{N(N^2-1)}\Sigma\left(k - \frac{N+1}{2}\right)(m_k - M). \qquad (5)$$

T[illegible]ECULAR ANNUAL DECREASE OF NUMBER OF STILLBIRTHS IN DENMARK.

$s = 70{,}000$, $N = 25$, $M = 1{,}735$

Year.	k.	m_k.	$m_k - M$.	$k - \frac{N+1}{2}$.	$\left(k - \frac{N+1}{2}\right)(m_k - M)$.
1888	1	1,861	+126	−12	− 1,512
89	2	1,924	+189	−11	− 2,079
1890	3	1,830	+105	−10	− 1,050
91	4	1,779	+ 44	− 9	− 396
92	5	1,811	+ 76	− 8	− 808
93	6	1,788	+ 53	− 7	− 371
94	7	1,719	− 16	− 6	+ 96
1895	8	1,753	+ 18	− 5	− 90
96	9	1,714	− 21	− 4	+ 84
97	10	1,811	+ 76	− 3	− 228
98	11	1,797	+ 62	− 2	− 124
99	12	1,737	+ 2	− 1	− 2
1900	13	1,696	− 39	0	0
01	14	1,732	− 3	+ 1	− 3
02	15	1,694	− 41	+ 2	− 82
03	16	1,685	− 50	+ 3	− 150
04	17	1,682	− 53	+ 4	− 212
1905	18	1,705	− 30	+ 5	− 150
06	19	1,602	−115	+ 6	− 690
07	20	1,723	− 12	+ 7	− 84
08	21	1,694	− 41	+ 8	− 328
09	22	1,665	− 70	+ 9	− 630
1910	23	1,658	− 77	+10	− 770
11	24	1,658	− 77	+11	− 847
1912	25	1,638	− 97	+12	− 1,164
				Sum:	−11,590

As an example illustrating secular fluctuations I take the previously discussed series of stillbirths in Denmark.

We have in this case

$$\frac{N(N^2 - 1)}{12} = 1{,}300,$$

hence:

$$s\beta = -\ 11{,}590 : 1{,}300 = -\ 8.92.$$

From this we may draw the conclusion that the number of still-births in Denmark pr. 70,000 births per annum on the average is decreased by 8.92.

If the fluctuations are of an essential secular character we may write

$$m = M + (k - 13)(-\ 8.92)$$

as the number of stillbirths pr. annum. Apart from accidental fluctuations due to sampling we should therefore obtain a nearly normal series for the 25-year period if we calculated the number of stillbirths each year according to the expression: $m_k - (k - 13)(-\ 8.92)$. Such a computation is given below:

NUMBER OF STILLBIRTHS IN DENMARK FREED FROM SECULAR FLUCTUATIONS.
$s = 70{,}000,\ N = 25.$

Year.	k.	$m_k - (k - 13)(-\ 8.92)$.	Year.	k.	$m_k - (k - 13)(-\ 8.92)$.
1888	1	1,754	1900	13	1,696
89	2	1,826	01	14	1,741
1890	3	1,741	02	15	1,712
91	4	1,699	03	16	1,712
92	5	1,740	04	17	1,718
93	6	1,726	1905	18	1,730
94	7	1,666	06	19	1,675
1895	8	1,708	07	20	1,875
96	9	1,678	08	21	1,765
97	10	1,784	09	22	1,745
98	11	1,779	1910	23	1,747
99	12	1,728	11	24	1,756
			1912	25	1,745

A computation of the characteristics of this series gives:

$$M = 1{,}735, \quad \sigma = 37.09, \quad \sigma_B = 41.6, \quad 100\rho \text{ imaginary.}$$

The dispersion is now slightly subnormal and the coefficient of disturbancy is imaginary whereas in the original series in § 88 it had a value equal to 3.4.

92. Cancer Statistics.—Mr. F. L. Hoffman in his treatise "The Mortality from Cancer Throughout the World" gives some very interesting statistics on mortality from cancer in various localities. Through the kindness of Mr. Hoffman I am able to submit the following series relating to cancer among males in the City of New York (Manhattan and Bronx Boroughs):

Deaths from Cancer (m_k) in the City of New York as Reduced to a Stationary Population of 1,000,000.

$s = 1{,}000{,}000,\ N = 25,\ M = 560.$

Year.	k.	m_k.	$m_k - M$.	$k - \frac{N+1}{2}$.	$\left(k - \frac{N+1}{2}\right)(m_k - M)$.
1889	1	377	−183	−12	2,196
1890	2	476	− 84	−11	924
91	3	410	−150	−10	1,500
92	4	444	−116	− 9	1,044
93	5	462	− 98	− 8	784
94	6	423	−137	− 7	959
1895	7	442	−118	− 6	708
96	8	493	− 67	− 5	335
97	9	505	− 55	− 4	220
98	10	515	− 45	− 3	135
99	11	513	− 47	− 2	94
1900	12	547	− 13	− 1	13
01	13	595	+ 35	0	0
02	14	540	− 20	+ 1	−20
03	15	580	+ 20	+ 2	40
04	16	609	+ 49	+ 3	147
1905	17	639	+ 79	+ 4	316
06	18	619	+ 59	+ 5	295
07	19	658	+ 98	+ 6	588
08	20	631	+ 71	+ 7	497
09	21	683	+123	+ 8	984
1910	22	710	+150	+ 9	1,350
11	23	710	+150	+10	1,500
12	24	721	+161	+11	1,771
1913	25	718	+158	+12	1,896
				Sum: Σ =	18,276

A computation of the dispersion and the Charlier coefficient of disturbancy gives a value of 100ρ in the neighborhood of 18, indicating marked fluctuations. An inspection of the series shows immediately that there is a marked increase in the rate of death from cancer. Working out the secular disturbances in the ordi-

nary manner we find:

$$s\beta = \frac{18,276}{1,300} = 14.06$$

indicating an increase of death from cancer of about 14 persons pr. annum for a population of 1,000,000. Eliminating the secular disturbances in the same manner as above, we now get a coefficient of disturbancy equal to $0.983i$ ($i = \sqrt{-1}$), practically a normal dispersion when taking into account the mean error due to sampling.

93. Application of the Lexian Dispersion Theory in Actuarial Theory. Conclusion.—The Russian actuary, Jastremsky, has applied the Lexian Dispersion Theory in testing the influence of medical selection in life assurance.[1] The research by Jastremsky evolves about the following question. Is medical selection a phenomena independent of the age of the assured? Let ${}^{(t)}q_x$ denote the observed rate of mortality after t years' duration of assurance. In the same manner $q_x^{(5)}$ denotes the rate of mortality of a life aged x after 5 *or more* years of duration ($t \leqq 5$). Forming the ratio ${}^{(t)}q_x : q_x^{(5)}$ for all ages of x we obtain a certain homograde series for which we may compute the Lexian Ratio and the Charlier Coefficient and thus determine if the fluctuations are due to sampling only or dependent on the age of the assured. Space does not allow us to give a detailed account of the very interesting research by Jastremsky as applied to the Austro-Hungarian Mortality Table (Vienna, 1909), and we shall limit ourselves to quote his final results as to the Lexian Ratio, L, for Whole Life Assurances and Endowment Assurances:

	Whole Life Assurances.	Endowment Assurances
t	L	L
1	0.88	1.01
2	0.89	0.96
3	1.12	1.05
4	1.05	0.98
5	1.07	0.91

The above values of L all lie close to unity and the series may therefore be considered as a Bernoullian Series where the fluctu-

[1] Jastremsky, "Der Auslese-Koeffizient," *Zeitschr. f. d. ges. Vers.-Wiss.*, Band XII 1912.

ations are due to sampling entirely. Or in other words, the ratio $\varphi_t = {}^{(t)}q_x : q_x^{(5)}$ is a quantity independent of the age of the assured.

The great majority of statistical series may be subjected to a similar analysis as given in the preceding chapters. The characteristics as described previously, the Lexian Ratio and the coefficient of disturbancy, tell us the magnitude of possible fluctuations from sample to sample. In many cases we may by means of the secular coefficient of disturbancy, β, partly or wholly eliminate such fluctuations, due to secular causes, and thus be in a better position to study the periodical fluctuations.

A statistical research may be likened to the navigation of a difficult waterway, full of hidden rocks and skerries out of sight to the navigator. The amateur statistician, sailing the ocean in a blind and happy-go-lucky manner, often comes to grief on those rocks and suffers a total shipwreck. The skillful navigator, the mathematically trained statistician, is always on the lookout for the sea marks. In the Lexian Ratio and the Charlier Coefficient of Disturbancy he recognizes a beacon light, often signalling "Danger ahead." He stops his engines. In case he does not possess the particular charts giving the exact location of the hidden reefs his prudence advises him to call a pilot to bring his ship safely in harbor. On the other hand, if he has reliable charts and knows his profession thoroughly he may venture forth and do his own piloting, by a study of the charts. It is to the study of such charts—*i. e.*, a special study of the higher statistical characteristics—that we shall turn our attention in the second part of this treatise. The reader who has followed us up to this point may perhaps feel discouraged by realizing how little he has gained in knowledge after having learned a mass of technical detail and formulas. We can quite appreciate and understand this feeling. So far, he has perhaps chiefly been impressed by the treacherous and misleading character of statistical mass phenomena, but to recognize a danger signal and thus avoid the pitfalls is one of the fundamental essentials in safe navigation in statistical research.

PART II

FREQUENCY CURVES AND HETEROGRADE STATISTICS

CHAPTER XIII.

THE THEORY OF ERRORS AND FREQUENCY CURVES AND ITS APPLICATION TO STATISTICAL SERIES. HISTORICAL NOTES.

94. General Remarks. The Hypothesis of Elementary Errors.—In the previous chapters we have discussed the elementary statistical parameters, the mean and the dispersion, together with the Lexian ratio and the Charlier coefficient of disturbancy and their application to the mathematical analysis of the homograde series. We shall now extend this discussion to the parameters of higher orders, such as the skewness and the excess, and also give the theory for a mathematical analysis of the other great domain of statistical series, the heterograde series.

The main reason for the separate treatment of the homograde statistical series is on account of their close analogy to ordinary mathematical probabilities. Whenever the *number of comparison*, s, may be regarded as equivalent to the total number of equally possible cases in ordinary a priori probabilities and the observed occurrences of the attribute (event) as the favorable number of cases, m, among the total number of possible events, s, we are justified in regarding the ratio $m : s$ in the light of a mathematical probability. For this reason all homograde series may be explained as being subject to the same mathematical laws as those governing ordinary a priori probabilities, which are fully explained by the series of Bernoulli, Poisson and Lexis and the various combinations of such series. Moreover, in all such series it is possible to compute both the mean and the dispersion by the indirect or combinatorial process instead of the direct or physical process.

The nucleus of the three fundamental series, the Bernoullian, the Poisson and the Lexian, as well as their various combinations is found in the development of the point binomial $(p + q)^s$ of the Bernoullian Theorem as described in Chapter IX where the general term expressing the probability of the occurrence of an event E α times and of the complementary event $\overline{E}$, $\beta = s - \alpha$ times is given by the formula

$$\varphi(\alpha) = \binom{s}{\alpha} p^{\alpha} q^{s-\alpha}$$

The numerical computation of this exact expression becomes, however, too unwieldy for large values of s and we shall therefore try to replace it with a more flexible approximation, preferably by a continuous function or by a rapidly convergent infinite series. On page 101 we gave such an approximation formula for the maximum value of φ (α), denoted by the symbol T_m. We wish, however, to find a simpler expression for the more general term as well. This further development necessitates the determination of several higher characteristics or parameters than those expressed by the mean and the dispersion. If we should succeed in this task the homograde series can be fully explained by the theory of mathematical probabilities and placed upon a sound a priori basis.

The question now arises whether the a priori theory of mathematical probabilities will furnish a similar basis for the second domain of statistics, the heterograde series. We are of course able to compute by means of the direct or physical process both the mean and the dispersion in various heterograde series, such as measurements on heights of adult males, number of fin rays in fishes or number of telephone calls over a trunk line in a given interval of time. But are we also able, like in the case of the homograde series, to forecast those parameters by means of the criterions of the Bernoullian, Poisson and Lexian series, i. e. by the indirect or combinatorial process?

A simple consideration will soon lead us to the admission that no a priori reasoning or a simple theorem like that of the Bernoullian will enable us to forecast the mean stature of Danish, Norwegian, Swedish or English adult males or the mean number of telephone calls over a trunk wire in a given interval of time. And while we by the physical or direct process can compute both the mean and the dispersion from previously collected statistical data, we have no way of knowing whether such parameters, purely empirical in form and nature, have any real significance beyond that of abstract mathematical calculations. Nor do such empirical parameters offer similar explanations as those of the homograde series. We are for instance able to predict the probability that in a series of 1000 successive drawings (with replacements) from a deck of whist the number of drawn aces will fall between 100 and 120. But we are not able by an a priori reasoning or by mathematical deduction to forecast the probability that among 1000 Scandinavian adult males—all chosen at random—the height

of an arbitrarily selected individual will fall between 170–175 centimeters.

Experience has, however, shown that the heterograde series show similar grouping tendencies around the mean value as those encountered in the homograde series. As an example we may compare the Bernoullian series of black cards in sample sets of 10 as collected by M. Charlier and shown on page 138 and the Poisson series of black balls in sample sets of 20 collected by Mr. Bonynge (shown on page 143) with a series of measurements, relating to the heights of Danish conscripts for the year 1916. Such a comparison is given below.

Charlier's Data			*Bonynge's Data*			*Danish Anthropometric Data*		
m	x	$F(x)$	m	x	$F(x)$	w	x	$F(x)$
0	−5	3	0	−5	2	140–145 cm.	−5	32
1	−4	10	1	−4	9	146–150 "	−4	44
2	−3	43	2	−3	35	151–155 "	−3	243
3	−2	116	3	−2	52	156–160 "	−2	1284
4	−1	221	4	−1	86	161–165 "	−1	3777
5	0	247	5	0	109	166–170 "	0	5742
6	+1	202	6	+1	85	171–175 "	+1	4796
7	+2	115	7	+2	69	176–180 "	+2	2129
8	+3	34	8	+3	30	181–185 "	+3	588
9	+4	9	9	+4	16	186–190 "	+4	81
10	+5	0	10	+5	6	over 191 "	+5	11
			11	+6	1			
		1000			500			18727

The grouping tendency or the clustering around the mean value is manifest in all three series; but while this tendency in the case of the two homograde series as offered in the experiments by Charlier and Bonynge may be fully explained by means of the theorems of mathematical probabilities no such reasoning is sufficient to explain the clustering tendency of the heterograde series relating to Danish conscripts. The calculus of probabilities in itself would not be sufficient to explain the grouping tendency of the variates in a heterograde series unless a general hypothesis will aid us in explaining the variation among several heterograde objects in respect to a specific attribute.

Thus the question which now confronts us is whether it is possible to establish a simple hypothesis which will enable us to extend the principles of the mathematical theory of probabilities

to the domain of the heterograde series and to build up a theory similar to that of the homograde series. The great Laplace was the first to solve this problem, and his investigations and analysis on this important subject are indeed some of the most important, but also some of the most difficult to follow in his *Théorie des Probabilités*. The hypothesis employed by Laplace in explaining the phenomena of variation in a heterograde series is the *hypothesis of elementary errors*. The hypothesis was later on somewhat simplified by the German astronomer and engineer, Hagen, and it has of late years been further developed through the elegant researches of the Scandinavian astronomer and statistician, M. Charlier. According to the Laplacean—Hagen—Charlier theory *every variate or individual deviation from a certain norm is generated as the sum of a mass of small and unknown quantities—generally infinite in number—which are known as elementary errors (deviations)*. The word error must of course be taken in a different sense than that we usually associate with the word. In precision measurements we are actually dealing with true or natural errors arising from imperfections of the instruments and the observer, but it would of course not be right to regard a deviation of say 5 centimeters from the mean stature of a population group as an error in the usual sense of the word. Used in its wider sense as an expression for deviations the term will, however, be readily understood, and it is in this sense we shall use it in the following pages.

Expressed in mathematical symbols the hypothesis of elementary errors may be presented as follows. Let x_k (where $k = 1, 2, 3, \ldots s$ denotes the kth error source among a total of s sources) represent the magnitude of a statistical variate expressed as a deviation (error) from a certain norm, then

$$f_k(r) \ (k = 1, 2, 3, \ldots s)$$

may be regarded as the probability that x_k assumes the value r. As to this particular elementary error probability function Laplace makes no other assumptions than those which follow directly from the definition of a mathematical probability. That is to say

$$0 \leqq f_k(r) \leqq 1,$$

where $k = 1, 2, 3, \ldots s$ and $r = 0 \pm 1, \pm 2, \pm 3, \ldots \pm \infty$

Since it is certain that one or more of the above values of r,

whether positive or negative, are bound to occur, we have evidently the relation

$$\sum_{r=-\infty}^{r=+\infty} f_k(r) = 1 \ (k = 1, 2, 3, \ldots s).$$

The epoch-making analysis of Laplace lies in the determination of the unknown function $f_k(r)$ from such simple and general assumptions.

95. Application to Statistical Series. Definitions.—The Laplacean-Charlier hypothesis of elementary errors opens the way for a mathematical analysis of a vast number of statistical data and series, which we shall briefly discuss in the following paragraph. First of all we submit therefore the following definition of a *statistical object.*

A number of similar objects (a species) which can be arranged in numerical order according to the measurable variation of a certain observed attribute (character), also called a variate, is known as a statistical object, eventually as a statistical series.

It is readily seen that this definition covers a wide range of subjects and that statistical methods instead of being applicable to social and economic problems only are equally useful in botany, zoölogy, biology and even in astronomy, physics or chemistry. Moreover, since the deviation of an individual variate of the statistical series as measured from a certain arbitrarily choosen norm evidently may be regarded as the sum of several elementary errors (the word error to be taken in its wider sense), it is evident that the statistical object can be subjected to a mathematical analysis on the basis of the theory of errors.

A simple consideration will also convince the reader that the above definition covers not alone the heterograde series but also the homograde series. For instance in the Bernoullian and Poisson series as presented in the experiments by Charlier and Bonynge on page 170, the number m, which gives the number of favorable events in each sample set, may be considered as a statistical variate and $F(x)$ as a statistical series.

This simple fact is of the utmost importance since it makes it possible to treat both the homograde and heterograde series on the common basis of elementary errors and links in the case of the homograde series the a priori mathematical probabilities with the a posteriori probabilities. Such connecti

is of special interest in the further treatment of the celebrated Rule of Bayes.

While thus the homograde and heterograde series may be viewed from a common viewpoint it is, however, necessary to point out a distinct difference in the nature of the statistical variates themselves. In one case we find the variate (the measurable attribute) expressed in whole numbers only, such as the number of fin rays in fishes, petals in flowers or the occurrence of a specified color in card drawings. The variates are in such cases known as *integral variates.* The observations on tail fin rays of flounders by the Danish biological station on page 131 offer such an example. As a further illustration of integral variates we choose the following statistical series from the observations of the English physicists, Rutherford and Geiger. Messrs. Rutherford and Geiger counted the numbers of alpha particles radiated from a bar of polonium during a long series of intervals, each lasting one-eighth of a minute. The table states the number of times, $F(x)$, the number of particles omitted in this interval had a given value, x.

x	$F(x)$
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	0
13	1
14	1

As an example of a very slight variation the Danish biologist and botanist, W. Johannsen, quotes the following observations by his colleague Professor Raunkjœr of Copenhagen on the number of involucral leaves of 100 samples chosen at random of *taraxacum erythrospernum:*

No. of Leaves x	*Frequency* $F(x)$
13	99
14	1

In other cases it is not possible to express the measure of the attribute in whole numbers. Thus measurements of stature, chest circumference and weight of recruits, or measurements of the percentages of sterility in wheat, barley, rye and oats will in general possess all possible fractional values between two integral numbers. Hence we must group the observations in classes, and such classified variates are known as *graduated variates.*

The measurements of heights of Danish conscripts for the year 1916 and shown on page 170 offer an illustration of graduated variates. Another case is furnished in the number of deaths by attained ages in a mortality table. In most mortality tables the deaths are given by integral ages only and represent therefore strictly speaking integral variates.

In biology we encounter numerous homograde series especially in investigations on dimorphism or polymorphism. Johannsen of Copenhagen produced from crossbreeding between a species of beans with white blossoms and yellow seeds and a species with violet blossoms and black seeds a bastard species with violet blossoms and muddy colored seeds. The offsprings—558 individual plants—of this bastard showed following variations:

White Blossoms		Violet Blossoms	
160		398	
Color of Seeds		Color of Seeds	
yellow	bronze	violet	black
39	121	105	293

In respect to blossoms we have two alternatives, in respect to seeds 4 alternatives.

As a few illustrations of the wide range of variable phenomena which allow to be classified as statistical objects, we present the following table:

Statistical Object.	*Sample.*	*Attribute.*	*Variate.*	*Series.*
Drawings of diamonds from deck of card.	Series of s drawings.	Frequency of Diamonds.	Integral.	Homograde.
Petals of Flowers.	Individual flowers.	No. of Petals.	"	Heterograde.
Fin Rays of Fish.	Individual fishes.	No. of Rays.	"	"
Sex at Birth.	Series of s births.	Frequency of Male Births.	"	Homograde
Age Distribution of Assured Persons.	The Policy-holder.	Age.	Graduated.	Heterograde.
Distribution of Amounts Assured.	The Policy.	Amount in Dollars.	"	"
Antropometric Measures of Recruits.	The Recruit.	Chest Measure.	"	"
Precision Measures.	Individual Measure.	Error.	"	"
Invalidity.	The Individual Patient.	Duration in days of Invalidity.	"	"
Income and Wages.	The Individual Worker.	Yearly Income in Dollars.	"	"
Cross of Flowers.	Series of s Bastard Flowers.	Color of Blossoms, Seeds, etc.	Integral.	Homograde.

It is to be noted that in the primary observations on homograde series the numbers are all abstract, whereas the heterograde series consist of concrete numbers. Another peculiarity of the homograde series is that they are always connected with the number of comparison, s, which is absent in the heterograde series.

96. Compound Frequency Curves.—According to the Laplacean-Charlier hypothesis any frequency curve may be considered as being generated as a sum of independent frequency curves and represents therefore in the final

instance really a compound frequency curve. Mathematically speaking we may therefore consider any frequency curve, no matter of what form, as being represented by the symbolic relation.

$$F(x) = \sum N_i \varphi_i(x) \qquad \text{for } i = 1, 2, 3, \ldots$$

The functions $\varphi_i(x)$ may sometimes all be normal or Laplacean probability curves. On the other hand, by assigning different values to N_i or the areas of the separate curve we may obtain a compound curve of wavelike form. Suppose for instance we had two samples of observations on the heights of say 100,000 Japanese recruits and 10,000 Danish recruits, each individual's measure written on separate cards. Suppose furthermore that we mixed those 110,000 cards in an urn, and then formed a new frequency distribution of this mixture This new frequency distribution would be a compound curve with two strongly pronounced crests or maximums. One (the Japanese) clustering around the value of 160 centimeters, while a smaller crest (the Danish) would tend to cluster around the value of 170 centimeters.

Another instance is offered in the distribution of the frontal breadths of a Naples specimen of the crab, *Carcinus mœnas*, as measured by Weldon. Weldon thought it very probable that this rather skew frequency distribution was produced by a fusion of two distinct races or species of individuals, which were clustered symmetrically around separate means. The distinguished English biometrician, Karl Pearson, tested this hypothesis for him and analysed the compound curve as two component curves representing respectively 58.55 and 41.45 per cent of the total area of the compound frequency curve. Thus Weldon's hypothesis was verified by a mathematical analysis.

A quite different type of example is offered in the frequency distribution of deaths by attained age as represented by the d_x column in any ordinary mortality table. The fact that the deaths in the d_x column showed a marked clustering tendency, strongly suggestive of the normal Laplacean curve around the age group 70–75, was already noted by Lexis, who in this way made a very interesting attempt to determine what he called a NORMALALTER for the age of man. Later on Italian statisticians took up the problem and analysed the d_x curve of Italian life tables as a sum of several normal frequency curves. Karl Pearson was the third investigator to take up the problem in a very fascinating essay in his *Chances of Death*. Pearson pictures Death as 5 marksmen shooting at a human target passing over the Bridge of Life. Each marksman aims with different precision and skewness. The result is 5 component skew curves.

Although the brilliant and perfect literary style of the eminent English biometrician rouses the admiration and brings forth the reader's unstinted praise, I can, however, not help being in accord with the distinguished American biologist and statistician, Raymond Pearl, who in his 1920 Lowell Lectures, although speaking in the highest terms of praise of Pearson's work, characterized a mathematical analysis of this kind as being nothing more than a highly interesting and neat graduation formula but wholly void of any biological significance.

It is as a mere matter of fact a comparatively easy matter to break up any death curve of a mortality table into separate mathematical components. As an example of such a process I offer the illustration in Chapte

I have broken up the recently published American $AM^{(5)}$ mortality table into two curves of the Gram-Charlier type, obtaining as good results as the Italian investigators and Pearson, who use 5 component curves.

But as already pointed out by Lexis "a mere mathematical analysis in component groups does not enlarge our knowledge of the causal relationships. It would be a quite different matter, however, if it were possible to establish clustering tendencies around definite ages for each of the more important causes of death."

An attempt to do this has been made by the present writer in his forthcoming book *An Elementary Treatise on Frequency Curves and their Application to the Human Death Curve*. I start with the hypothesis "that the frequency distribution of deaths at attained ages classified according to certain groups of causes of death among the survivors in a mortality table tend to cluster around specific ages in such a manner that their frequency distribution can be represented by a Gram-Charlier frequency curve." If this hypothesis can be accepted as having a sound biological basis I have shown that it is possible by a mathematical analysis resting on such hypothesis to construct mortality tables from mortuary records by sex, attained age and cause of death, and *without any information about the number of lives exposed to risk at various ages.* This proposal has been met by a storm of protest from many American actuaries, who claim that I have attempted the impossible. Final judgment should be suspended, however, until the actual appearance of the work, which I think must be judged from a biological rather than from a mathematical point of view. The fact that the method has given good results in the construction of many mortality tables among highly different races and occupations must, I think, be attributed to purely biological causes and not to actuarial or mathematical methods, which in the process have been employed as a mere tool, as a means rather than as an end.

97. Early Writers.—The idea of frequency curves or frequency distribution is probably very old. It very likely arose in the mind of man when he began to make quantitative observations. Undoubtedly the surveyors and engineers of the people of ancient civilization had noticed that successive and independent measurements of the same object often showed variations. On the other hand we have no means of knowing if the ancient geometers and mathematicians knew how to estimate and value such variations from the true value of the object. It is probable that the great Greek astronomers, such as Hipparch and Aristarch in their astronomical observations have employed some rational method of allowing for errors due to the instruments and the individual observer, but no records are available so as to settle this question.

The great Danish astronomer, Tycho Brahe, the father of modern astronomy, on the other hand made careful adjustments for errors of observations and has left us records on the systematic method of such adjustments.

However, it was not before the close of the eighteenth century that the errors of observations were subjected to a mathematical treatment. The first known writer on the mathematics of the subject was the English actuary, Thomas Simpson, a most remarkable self-taught mathematician who in 1757 issued his "Miscellaneous Tracts on some curious and very interesting subjects in Mechanics, Physical Astronomy and Speculative Mathematics." In this interesting and instructive little book is found a chapter entitled "An Attempt to shew the Advantage arising by Taking the Means of a Number of Observations in Practical Astronomy."

About 15 years later the French mathematician, Lagrange, took up the ideas of Simpson in a memoir, which at the time caused considerable notice in mathematical circles. Lagrange in his treatment followed a course very much similar to that employed by de Moivre in the discussion of the problem which bears his own name.

In 1778, Daniel Bernoulli in the scientific publications of the Russian academy of Petrograd subjected the memoirs of Lagrange to a searching criticism and proposed the first mathematical formula for a frequency curve or curve of errors around the mean. Bernoulli suggested as a law of error or frequency function, $\varphi(x)$, the following expression:

$$\varphi(x) = +\sqrt{r^2 - x^2}, \text{ where } r \text{ is a constant}$$

This equation represents a symmetrical semi-circle and gives as we shall have occasion to show at a later stage a rough approximation to the presumptive law of error.

A very important contribution to the theory was also made by the American, Adrain, in his journal "The Analyst."

98. Laplace and Gauss.—Laplace was the next mathematician to take up the subject of frequency curves in his monumental work "Théorie analytique des probabilités." The great Frenchman dealt with the subject in a manner which leaves little to be desired. M. Charlier, the eminent Swedish astronomer and statistician, has justly remarked that among the various deductions of the law of errors, the exhaustive researches of Laplace occupy beyond doubt a leading position because of their generality and far reaching applications. On the other hand, the analysis of Laplace is by no means easy to follow in all its details and the 4th chapter of the "Thêorie Analytique des Probabilitês" accord-

ing to a remark by Todhunter in his "History of Probabilities" forms one of the most important but at the same time also one of the most difficult parts of the great work.

No doubt the extreme difficulty of fully mastering the far reaching but intricate analysis of Laplace was realized by the mathematicians. Already his friend and disciple, Poisson, realized this and issued in 1832 a note entitled "Sur la probabilité des resultats moyens des observations." But the wealth of ideas in Laplace's treatise and their wide range of application were really never fully recognized for almost a full century when they were taken up by Charlier, who more than any one else has proven their great worth as the most general and direct basis for a complete theory of frequency functions and the associated problem of correlation.

In the meantime Laplace's method had been supplanted by the independent and contemporary researches of the great German mathematician, Gauss. The method employed by Gauss in deriving his law of error or frequency curve and the therewith associated criterions of the method of least squares is undoubtedly very simple and elegant and much easier to follow for the beginner than the analysis of Laplace. Gauss in his studies confided himself to the so-called precision errors or errors arriving from repeated measurements by means of physical instruments, such as astronomical or geodetic observations or measurements in experimental physics or chemistry.

The ideas put forth by Gauss were followed up by a number of astronomers and physicists, such as Bessel, Encke, Hansen, and Hagen of Germany, Andræ, D'Arrest, and Gylden of Scandinavia, Airy, Herschell, and Tait in England, Laurent in France, and Newcomb and Chauvenet in America. And the Gaussian methods are still used exclusively in preference to those introduced by Laplace in most of our text-books on theory of errors and the related subject of least squares.

One reason for this preference for the theory of Gauss apart from its simplicity of representation is to be looked for in the fact that until a comparatively recent date the majority of applications of the theory of frequency curves or error curves had reference to precision measurements. As pointed out by N. R. Jörgensen,[1] in his excellent Danish treatise on "Frequency Surfaces and Correlation" it will, as a rule, be found that the Gaussian error law may be

[1] Undersögelser over Frequensflader og Korrelation (Copenhagen, 1916).

regarded as an excellent method of approximation, which becomes well-nigh perfect in the case of errors of precision measurements with delicate instruments in the hands of carefully trained observers. The Gaussian frequency curve may therefore be said to fulfill all the requirements in praxis of a law of error, where we are concerned with errors in the true sense of the word.

99. Quetelet's Studies.—Matters became, however, quite different when the biologists and economists began to employ mathematical analysis in their research work. It was the great Belgian astronomer and statistician, Quetelet, who first introduced exact measurements in the study of biological and anthropological phenomena and showed that a number of collected statistical data on heights, weights and chest measurements of recruits exhibited a close conformity to the Gaussian law of error, although the variation among the individual objects as measured could not be considered solely as errors in the original sense of the word.

Investigations along this line were greatly accelerated by the discoveries of Quetelet. All sorts of measurements were taken and the rapidly growing collections of statistical data relating to economic and social conditions as recorded by various governmental statistical bureaus furnished material for further investigations. But unfortunately in all these investigations the Gaussian error law came to act as a veritable Procrustean bed to which all possible measurements should be made to fit. The belief in authority so typical of modern German learning and which had also spread to America was too great to question the supposed generality of the law discovered by the great Gauss. Statisticians could not conciliate themselves with the thought of the possible presence of "skew" frequency curves, although numerous data offered complete defiance to the Gaussian dogma and exhibited a markedly skew frequency distribution. Supposedly great authorities argued naïvely that the reason the data did not fit the curve of Gauss was that the observations were not numerous enough to eliminate the presence of skewness. In other words, skewness was regarded as a by-product of sampling and was believed could be made to disappear completely if we could take an infinite number of observations.

Voices had, however, been raised against these energetic but futile Procrustean efforts. Already Quetelet realized the existence of skew frequency curves. This is clearly brought out in his correspondence on this subject with Mr. Bravais of the École

Polytechnique of Paris as published in an appendix to his "Lettres sur la Théorie des probabilités."

100. Opperman, Gram, and Thiele.—Neither Quetelet nor Bravais succeeded, however, in giving a complete mathematical treatment of the theory of skew frequency curves. The first complete mathematical demonstration of this aspect of the matter was given by various Scandinavian investigators. A Danish actuary, Opperman, was probably the leading spirit in organizing the revolt against the belief in authority as preached by the adherents of the doctrine of Gauss. Opperman, who was a self-taught mathematician, seems to have looked with suspicion on many of the researches by German mathematicians of the latter half part of the nineteenth century. He was a great admirer of the early Scotch and English mathematicians with whose works he was thoroughly familiar, and it is said he took great delight in pointing out how many of the lengthy and formidable German demonstrations in the realm of the theory of functions had been demonstrated in a more elementary and clearer manner by such men as Wallis, Stirling, MacLaurin, Gregory, Briggs and Napier. As a practical actuary and managing director of the Danish Government Life Assurance Fund he had ample opportunity to notice that many frequency distributions occurring in actuarial work offered a notorious defiance to the frequency curve of Gauss. Around Opperman there gathered a small group of young enthusiastic students of mathematics among whom we may especially mention Gram and Thiele and to whom he expounded his ideas. Opperman himself wrote very little and always in a condensed form. A reviewer of his work remarks that he rewrote his essays several times so as to be able to represent on a single page what other mathematicians usually required a dozen of pages tc express. He has left very little material bearing on the theory of frequency curves, but his discussions on this subject with his younger disciples evidently bore fruit.

J. P. Gram was the first mathematician to show that the normal symmetrical Gaussian error curve was but a special case of a more general system of skew frequency curves which could be represented by a series. In his very original doctor's thesis in Danish on "The development of series by means of the method of least squares" (Copenhagan, 1879) [1] he extended some theories originally expounded by the Russian mathematician, Tchebycheff, to

[1] Om Rækkeudviklinger.

the representation of frequency functions by means of a series. By using the Gaussian curve

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}}$$

as a generating function Gram showed that an arbitrary frequency function could be represented approximately by a series of the form

$$F(x) = c_0\varphi(x) + c_1\varphi'(x) + c_2\varphi''(x) + c_3\varphi'''(x) + c_4\varphi''''(x) + \ldots$$

In this development Gram established some far reaching properties of infinite determinants and their relations to orthogonal functions which later have become of much use in the recent epochmaking researches on integral equations by the Swedish mathematician and actuary, Fredholm.

To Gram, therefore, belongs the honor of having been the first mathematician to give a systematic theory for the development of skew frequency curves.

While Gram's later work as a managing director of a life insurance company occupied most of his time and left but little opportunity for purely mathematical work his friend, T. N. Thiele, began to lecture at the University of Copenhagen on the general theory of observations. The substance of these decidedly original lectures was in 1889 published in book form under the title "A General Theory of Observations." [2]

In several respects this work occupies a dual position to the work of the great Laplace, although Thiele is set like flint against the idea of basing the theory of probabilities on the conception of an a priori probability. In his lectures he always maintained that the greatest benefit derived from the study of the method of least square was that the student learned where not to use it. Among one of the great achievments of Thiele is the introduction in the theory of frequency curves of a certain system of statistical characteristics to which he gave the name of *semi invariants* and which are practically identical to the system of moments later on introduced by Pearson. By means of these semi invariants Thiele arrived at the same series as deduced by Gram. In Thiele's work we also find a very original treatment of the theory of correlation

[2] Almindelig Iagttagelseslære.

originally introduced by Bravais. But instead of the term correlation he uses the words "bonded observations."

Like Laplace's "Théorie des Analytiques des Probabilités" Thiele's original work and its subsequent abridged translation in English offers by no means an easy reading, especially to the beginner. It contains, however, like the work of Laplace a veritable wealth of ideas and methods which remain unsurpassed in the realm of mathematical statistics and no serious worker on the general theory of observations can afford to neglect to study the original works of Laplace and Thiele.

101. Modern Investigations.—The investigations of Gram and Thiele bring us up to the close of the nineteenth century. Their ideas reached but a small number of students of mathematical statistics because of the very limited knowledge of Scandinavian languages among mathematical readers in general. But from the beginning of the nineties other voices began to be heard against the Gaussian dogma. In Germany it was Fechner who first entered the ranks of the opposition with his so-called "zweispaltiges Gesetz." His work was continued by Lipps and the Leipzig astronomer, Bruhns, who by the publication in 1906 of his "Kollektivmasslehre" gave an almost complete theory of frequency curves where we again find the series originally developed by Gram and Thiele.

Although quite considerable valuable work has thus been done in Germany along the lines of frequency curves it was, however, in England that the renewal of the classical probability theory took place with the renowned memoirs by the English mathematician, Karl Pearson, entitled "Contributions to the Mathematical Theory of Evolution" in *Philosophical Transactions* for 1895.

Since that year Pearson has produced a certain type of statistical literature, of almost Gargantuan proportions. The quarterly journal "Biometrika" of which he is the editor is devoted to the mathematical study of biological problems. When Pearson first introduced his famous types of curves (now more than 12 in number) the study of frequency distributions was greatly accelerated. The application of these curves to biological problems was apparently so simple that they were used in a rather loose manner by many biologists and anthropologists who had but little training in mathematical analysis. Examples of such pseudo mathematical analysis are especially found in the writings of the American

anthropologist, Franz Boas, which may be held up as a warning to all statisticians to keep away from the higher mathematical analysis of collected statistical data unless they are familiar with the tools of the probability calculus.

Such misuse can of course not be laid at the door of Mr. Pearson who indeed has protested vigorously against the erroneous application of his methods by investigators of the Boas type. On the other hand, it is equally true that Mr. Pearson at times has relied too much on his mathematical formulas and violated the maxim of the Danish biologist, Johannsen, that "we must practice biology *with* mathematics and not *as* mathematics."

The immense production of Pearson coupled with his well-nigh perfect and forceful style of writing has to a certain extent overshadowed the researches of his compatriot, Edgeworth, whose works according to the Danish actuary and mathematician, Jörgensen, are greatly superior to those of Pearson both in scientific rigor and in practical applications. Edgeworth has deduced the previously mentioned series by Gram and Thiele in a very elegant form in the *Cambridge Philosophical Transactions* (1904) and he has in a series of articles in the *Journal of the Royal Statistical Society* outdistanced many of his contemporaries among the mathematical statisticians. Unfortunately Edgeworth's contributions have not gotten the attention they deserve, probably because of the rather fragmentary and unsystematic manner in which they have appeared. Among the new methods introduced by Edgeworth we may particularly mention the so-called method of translation.

The Pearsonian types of frequency curves are represented by formulas which in mathematical language are termed closed expressions in contradistinction to the development in series. This latter method is still being preferred by the Scandinavian mathematicians under the leadership of the Swedish astronomer Charlier. Charlier started his first investigation with a small brochure entitled "Über das Fehlergesetz" in the *Meddelendan* for 1905, wherein he followed the method originally introduced by Laplace and in a most elegant way of deduction reached the series of Gram and Thiele. He has since that year published a series of small monographs on various aspects of mathematical statistics and their application to stellar statistics which beyond doubt are destined become classics in the history of probabilities.

Charlier has shown that all frequency curves fall into two t

which he has designated as type A and type B. The A type is the usual expansion of Gram and Thiele with the normal frequency curve as the generating function. Type B which covers decidedly skew frequency curves is given by the series

$$F(x) = c_0\,\varphi(x) + \frac{c_1}{\underline{|1}}\,\Delta\,\varphi(x) + \frac{c_2}{\underline{|2}}\,\Delta^2\,\varphi(x) + \ldots + \frac{(-1)^n}{\underline{|n}}\,\Delta^n\,\varphi(x) + \ldots$$

where

$$\varphi(x) = \frac{e^{-\lambda}}{\pi}\int_0^{\pi} e^{\lambda\cos\omega}\cos\,[\lambda\sin\omega - x\omega]\,d\omega$$

is the generating function.

The decidedly constructive work begun by Charlier has been ably supplemented by his talented disciple Wicksell and the Danish actuary, Jörgensen. Wicksell in 1920 issued in Swedish a series of lectures on mathematical statistics delivered during the autumn of 1919 before the Swedish Assurance Society. He has also written numerous excellent monographs on mathematical statistics and their application to vital statistics.

N. R. Jörgensen issued in 1916 his large octavo volume on "Researches on Frequency Surfaces and Correlation" [1] which beyond doubt is the most important work among the contributions of Danish actuaries since the appearance of the memoirs of Gram and Thiele. Jörgensen's systematic treatise has greatly furthered the studies of the Scandinavian school both in theoretical and practical aspects. A very important feature of his book is the insertion of an extensive collection of numerical tables of various functions which greatly facilitates the practical applications of the theory. These tables, many of which are the results of his individual efforts, hold equal rank with the well-known "Tables for Biometricians and Statisticians" edited by Karl Pearson in 1914.

Besides the writings of Charlier, Wicksell and Jörgensen a number of Scandinavian mathematicians, actuaries and statisticians have contributed valuable researches both on frequency curves and correlation methods. We may especially mention such men as Guldberg, Gyllemberg, Malmquist, Burrau and Lundquist. In this group we might also include the Danish biologist, Johann-

[1] Undersögelser over Frequensflader og Korrelation.

sen, whose writings on the theory of heredity are recognised as standard texts on the application of the mathematical statistical methods to problems dealing with inherited characteristics in organic life.

A very interesting attempt to develop a theory of frequency curves has been made by the Dutch astronomer, Kapteyn, in his "Skew Frequency Curves in Biology and Statistics" (Groningen, 1912). Kapteyn's theory which has much in common with Edgeworth's method of translation introduces a new idea in the generation of a frequency curve by making the size of the individual object depend not alone upon the sources influencing a collection of such individuals but also upon the size of the object at a previously given time t. This idea of introducing the time factor in the theory of probabilities is, however, more justly credited to the French mathematician, Bachelier, whose large treatise on probabilities of which the first volume appeared a few years ago has introduced some new thoughts regarding the conception of continuous probabilities which are bound to strongly influence the whole theory.

Before closing this necessarily brief and incomplete historical note we wish to mention the close connection of the theory of frequency curves with that of integral equations. Since the appearance of the epoch-making memoirs by Fredholm the theory of integral equations has occupied a central position in mathematical analysis. This youngest branch of higher analysis has already found numerous practical applications in physics and chemistry and possesses equally important properties in the way of solving numerous statistical problems. In fact, the whole theory of frequency curves and correlations can be reduced to the solution of a few integral equations whose constants contain all the characteristic properties of the frequency distribution. On the basis of this principle, a complete theory of frequency curves could be presented on a single book page.

CHAPTER XIV.

THE MATHEMATICAL THEORY OF FREQUENCY CURVES.

102. Frequency Distributions.—If N successive observations originating from the same essential circumstances or the same source of causes are made in respect to a certain statistical variate, x, and if the individual observations o_i ($i = 1, 2, 3, \ldots . N$) are permuted in their natural order in accordance with their magnitude then this particular permutation is said to form a frequency distribution of x and is denoted by the symbol $F(x)$.

The relative frequencies of this specific permutation, that is the ratio which each absolute frequency or group of frequencies bear to the total number of observations, is called a relative frequency function or probability function and is denoted by the symbol $\varphi(x)$.

If the statistical variate is continuous or a graduated variate, such as heights of soldiers, ages at death of assured lives, physical and astronomical precision measurements, etc., then

$$dz\,\varphi(z)$$

is the probability that the variate x satisfies the following relation

$$z - \frac{1}{2}dz < x < z + \frac{1}{2}dz$$

or that x falls between the above limits.

If the statistical variate assumes integral (discrete) values only as for instance the number of alpha particles discharged from certain radioactive metals and gases, such as polonium and helium, number of fin rays in fishes, or number of flower petals in plants, then $\varphi(z)$ is the probability that x assumes the value z. From the above definitions it follows directly that

$$\text{(a)} \quad F(z) = N\varphi(z) \quad \text{(Integral variates)}$$
$$\text{(b)} \quad dzF(z) = N\varphi(z)dz \quad \text{(Integrated variates)}$$

Interpreting the above results graphically we find that (a) will be represented by a series of disconnected or discrete points while (b) will be represented by a continuous curve.

As to the function $\varphi(z)$ we make for the present no other assumptions than those following immediately from the customary

definition of a mathematical probability. That is to say the function $\varphi(z)$ must be real and positive. Moreover, it must also satisfy the relation

$$\int_{-\infty}^{+\infty} \varphi(z)dz = 1,$$

or in the case discrete variates:

$$\sum_{z=-\infty}^{z=\infty} \varphi(z) = 1,$$

which is but the mathematical way of expressing the simple hypothetical disjunctive judgment that the variate is sure to assume some one or several values in the interval from $-\infty$ to $+\infty$. The zero point may be arbitrarily chosen and need not coincide with the natural zero of the number scale. Thus for instance if we in the case of Danish recruits choose the zero point of the frequency curve at 170 centimeters an observation of 180 centimeters would be recorded as $+10$ and an observation of 160 centimeters as -10.

103. Parameters considered as Symmetric Functions.—In regard to a frequency function we may assume a priori that it will depend only upon the variate x and certain mathematical relations into which this variate enters with a number of constants $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \ldots$, symbolically expressed by the notation

$$F(x, \lambda_1, \lambda_2, \lambda_3, \lambda_4 \ldots)$$

where the λ's are the constants and x the variate.

All these constants or parameters are naturally independent of x and represent some peculiar properties or characteristic essentials of the frequency function as expressed in the original observations o_i $(i = 1, 2, 3, \ldots N)$. We may, therefore, say that each constant or statistical parameter entering into the final mathematical form for the frequency function is a function of the observations o_i. This fact may be expressed in the following symbolic form

$$\lambda_1 = S_1(o_1, o_2, o_3, \ldots o_N)$$
$$\lambda_2 = S_2(o_1, o_2, o_3, \ldots o_N)$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\lambda_N = S_N(o_1, o_2, o_3, \ldots o_N)$$

But from purely a priori considerations we are able to tell something else about the function S_i. $(i = 1, 2, 3, \ldots N)$. It is only when permuting the various o's in an ascending magnitude according to the natural number scale that we obtain a frequency function. This arrangement itself has, however, no influence upon any one of the o's which were generated before this purely arbitrary permutation took place. The ultimate and previously measured effects of the causes as reflected in each individual numerical observation, o_i, depend only upon the origin of causes which form the fundamental basis for the statistical object under investigation and do not depend upon the order in which the individual o's occur in the series of observations.

Suppose for instance that the observations occurred in the following order

$$o_1, o_2, o_3, \ldots o_N$$

By permuting these elements in their natural order we obtain the frequency distribution $F(x)$. But the very same distribution could have been obtained if the observations had occurred in any other order as for instance

$$o_7, o_9, o_N, \ldots o_3 \ldots o_1.$$

so long as all of the individual o's were retained in the original records. Or to take a concrete example as the study of the number of policyholders according to attained ages in a life assurance office. We write the age of each individual policyholder on a small card. When all the ages have been written on individual cards they may be permuted according to attained age and the resulting series is a frequency function of the age x. We may now mix these cards just as we mix ordinary playing cards in a game of whist, and we get another permutation in general different from the order in which we originally recorded the ages on the cards. But this new permutation can equally well be used to produce the frequency function if we are only sure to retain all the cards and do not add any new cards.

The various functions $S(o_1, o_2, o_3, \ldots o_N)$ are, therefore, symmetric functions, that is functions which are left unaltered by arbitrarily permuting the N elements o, and no interchange whatever of the values of the various o's in those symmetric functions can have any influence upon the final form of the frequency function or frequency curve, $F(x)$.

We now introduce under the name of *power sums* a certain well-known form of fundamental symmetrical functions defined by the following relations

$$
\begin{aligned}
s_0 &= o_1^0 + o_2^0 + o_3^0 + \ldots o_N^0 = N \\
s_1 &= o_1^1 + o_2^1 + o_3^1 + \ldots o_N^1 = \Sigma o_i^1 \\
s_2 &= o_1^2 + o_2^2 + o_3^2 + \ldots o_N^2 = \Sigma o_i^2 \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
s_N &= o_1^N + o_2^N + o_3^N + \ldots o_N^N = \Sigma o_i^N
\end{aligned}
$$

Moreover, a well-known theorem in elementary algebra tells us that every symmetric function may be expressed as a function of $s_1, s_2, s_3, \ldots s_N$.

From this theorem it follows a fortiori that we are able to express the constants λ in the frequency curve as functions of the power sums of the observations. While such a procedure is possible, theoretically at least, we should, however, in most cases find it a very tedious and laborious task in actual practice. It, therefore, remains to be seen whether it is possible to transform these symmetrical functions of the power sums of the observations into some other symmetrical functions, which are more flexible and workable in practical computations and which can be expressed in terms of the various values of s.

104. Semi-Invariants of Thiele.—It is the great achievement of Thiele to have been the first mathematician to realize this possibility and make such a transformation by introducing into the theory of frequency curves a peculiar system of symmetrical functions which he called *semi-invariants* and denoted by the symbols $\lambda_1, \lambda_2, \lambda_3 \ldots$

Starting with the power sums, s_i, Thiele defines these by the following identity

$$s_0 e^{\frac{\lambda_1 \omega}{\lfloor 1} + \frac{\lambda_2 \omega^2}{\lfloor 2} + \frac{\lambda_3 \omega^3}{\lfloor 3} + \cdots} = s_0 + \frac{s_1 \omega}{\lfloor 1} + \frac{s_2 \omega^2}{\lfloor 2} + \frac{s_3 \omega^3}{\lfloor 3} + \ldots \qquad (1)$$

which is supposed identical in respect to ω.

Since $s_i = \Sigma o^i$ the right hand side of the equation may also be written as $e^{o_1 \omega} + e^{o_2 \omega} + e^{o_3 \omega} + \ldots = \Sigma e^{o_i \omega}$.

Differentiating (1) with respect to ω we have

$$s_0 e^{\frac{\lambda_1\omega}{\underline{|1}}+\frac{\lambda_2\omega^2}{\underline{|2}}+\frac{\lambda_3\omega^3}{\underline{|3}}+\dots}\left[\lambda_1+\frac{\lambda_2\omega}{\underline{|1}}+\frac{\lambda_3\omega^2}{\underline{|2}}+\dots\right]=$$

$$=\left[s_0+\frac{s_1}{\underline{|1}}\omega+\frac{s_2}{\underline{|2}}\omega^2+\frac{s_3}{\underline{|3}}\omega^3+\dots\right]\left[\lambda_1+\frac{\lambda_2\omega}{\underline{|1}}+\frac{\lambda_3}{\underline{|2}}\omega^2+\dots\right]=$$

$$=s_1+\frac{s_2}{\underline{|1}}\omega+\frac{s_3}{\underline{|2}}\omega^2+\frac{s_4}{\underline{|3}}\omega^3+\dots$$

Multiplying out and equating the various coefficients of equal powers of ω we finally have:—

$$\begin{aligned}
s_1 &= \lambda_1 s_0\\
s_2 &= \lambda_1 s_1 + \lambda_2 s_0\\
s_3 &= \lambda_1 s_2 + 2\lambda_2 s_1 + \lambda_3 s_0\\
s_4 &= \lambda_1 s_3 + 3\lambda_2 s_2 + 3\lambda_3 s_1 + \lambda_4 s_0\\
&\cdots\cdots\cdots\\
&\cdots\cdots\cdots
\end{aligned}$$

where the coefficients follow the law of the binomial theorem. Solving for λ we have

$$\begin{aligned}
\lambda_1 &= s_1 : s_0\\
\lambda_2 &= (s_2 s_0 - s_1^2) : s_0^2\\
\lambda_3 &= (s_3 s_0^2 - 3s_2 s_1 s_0 + 2s_1^3) : s_0^3\\
\lambda_4 &= (s_4 s_0^3 - 4s_3 s_1 s_0^2 - 3s_2^2 s_0^2 + 12 s_2 s_1^2 s_0 - 6s_1^4) : s_0^4\\
&\cdots\cdots\cdots\\
&\cdots\cdots\cdots
\end{aligned}$$

The semi-invariants λ in respect to an arbitrary origin and unit are defined by the relation

$$s_0 e^{\frac{\lambda_1\omega}{\underline{|1}}+\frac{\lambda_2\omega^2}{\underline{|2}}+\frac{\lambda_3\omega^3}{\underline{|3}}+\dots} = e^{o_1\omega}+e^{o_2\omega}+e^{o_3\omega}+\dots.$$

where o_1, o_2, o_3 . . . are the individual observations.

Let us now change to another coördinate system with another unit and origin defined by the following linear transformation

$$o_i' = ao_i + c$$

The semi-invariants in this new system are given by the relation

$$s_0 e^{\frac{\lambda_1'\omega}{\underline{|1}}+\frac{\lambda_2'\omega^2}{\underline{|2}}+\frac{\lambda_3'\omega^3}{\underline{|3}}+\dots} = e^{o_1'\omega}+e^{o_2'\omega}+\dots.$$

$$= e^{(ao_1+c)\omega}+e^{(ao_2+c)\omega}+\dots.$$

Since the various values of λ' do not depend upon the quantity ω we may without changing the value of the semi-invariants replace ω by $\omega : a$ in the above equations which give

$$s_0 e^{\frac{\lambda_1' \omega}{a\underline{|1}} + \frac{\lambda_2' \omega^2}{a^2\underline{|2}} + \frac{\lambda_3' \omega^3}{a^3\underline{|3}} + \cdots} = e^{(a o_1 + c)\frac{\omega}{a}} + e^{(a o_2 + c)\frac{\omega}{a}} + e^{(a o_3 + c)\frac{\omega}{a}} + \cdots$$

$$= e^{\frac{c\omega}{a}}\left[e^{o_1\omega} + e^{o_2\omega} + e^{o_3\omega} + \cdots\right] = e^{\frac{c\omega}{a}}\, s_0 e^{\frac{\lambda_1\omega}{\underline{|1}} + \frac{\lambda_2\omega^2}{\underline{|2}} + \frac{\lambda_3\omega^3}{\underline{|3}} + \cdots}$$

Taking the logarithms on both sides of the equation we have

$$\frac{\lambda_1'\omega}{a\underline{|1}} + \frac{\lambda_2'\omega^2}{a^2\underline{|2}} + \frac{\lambda_3'\omega^3}{a^3\underline{|3}} + \cdots = \frac{c\omega}{a} + \frac{\lambda_1\omega}{\underline{|1}} + \frac{\lambda_2\omega^2}{\underline{|2}} + \frac{\lambda_3\omega^3}{\underline{|3}} + \cdots$$

Differentiating successively with respect to ω we have

$$\frac{\lambda_1'}{\underline{|1}\, a} + \frac{\lambda_2'\omega}{a^2} + \frac{\lambda_3'\omega^2}{\underline{|2}\, a^3} + \cdots = \frac{c}{a} + \lambda_1 + \frac{\lambda_2\omega}{\underline{|1}} + \frac{\lambda_3\omega^2}{\underline{|2}} + \cdots$$

$$\frac{\lambda_2'}{a^2} + \frac{\lambda_3'\omega}{a^3} + \frac{\lambda_4'\omega^2}{\underline{|2}\, a^4} + \cdots = \lambda_2 + \frac{\lambda_4\omega^2}{\underline{|2}} + \cdots$$

$$\frac{\lambda_3'}{a^3} + \frac{\lambda_4'\omega}{a^4} + \cdots = \lambda_3 + \lambda_4'\omega + \cdots$$

. . . .

. . . .

Letting $\omega = 0$ we therefore have

$$\frac{\lambda_1'}{a} = \frac{c}{a} + \lambda_1, \text{ or } \lambda_1' = a\lambda_1 + c$$

$$\frac{\lambda_2'}{a^2} = \lambda_2, \text{ or } \lambda_2' = a^2\lambda_2$$

$$\frac{\lambda_3'}{a^3} = \lambda_3, \text{ or } \lambda_3' = a^3\lambda_3$$

. . . .

. . . .

from which we deduce the following relations

$$\lambda_1(ax + c) = a\lambda_1(x) + c$$
$$\lambda_r(ax + c) = a^r\lambda_r(x) \qquad \text{for } r > 1$$

We shall for the present leave the semi-invariants and only ask the reader to bear in mind the above relations between λ and s,

of which we shall later on make use in determining the constants in the frequency curve $\varphi(x)$.

Before discussing the generation of the total frequency curve it will, however, be necessary to demonstrate some auxiliary mathematical formulæ from the theory of definite integrals and integral equations which will be of use in the following discussion as mathematical tools with which to attack the collected statistical data or the numerical observations.

105. The Fourier Integral Equation.—One of these tools is found in the celebrated integral theorem of Fourier, which was the first integral equation to be successfully treated. We shall in the following demonstration adhere to the elegant and simple solution by M. Charlier. Charlier in his proof supposes that a function, $F(\omega)$, is defined through the following convergent series.

$$F(\omega) = \alpha\,[f(o) + f(\alpha)e^{\alpha\omega i} + f(2\alpha)e^{2\alpha\omega i} + \dots + f(-\alpha)e^{-\alpha\omega i} + f(-2\alpha)e^{-2\alpha\omega i} + \dots]$$

or

$$F(\omega) = \alpha \sum_{m=-\infty}^{m=\infty} f(\alpha m)e^{\alpha m\omega i} \tag{2}$$

where $i = \sqrt{-1}$

We then see by a well known theorem of Cauchy that the integral

$$I(\omega) = \int_{-\infty}^{+\infty} f(x)e^{x\omega i}dx \,[1] \tag{3}$$

is finite and convergent. If we now let $m\alpha = x$ and let $\alpha = 0$ as a limiting value, α becomes equal to dx and $f(\alpha m) = f(x)$. Consequently we may write

$$\lim_{\alpha=0} F(\omega) = I(\omega).$$

Multiplying (2) by $e^{-r\alpha\omega i}d\omega$ and integrating between the limits $-\pi/\alpha$ and $+\pi/\alpha$ we get on the left an expression of the form $\int_{-\pi/\alpha}^{+\pi/\alpha} F(\omega)e^{-r\alpha\omega i}d\omega$ and on the right a sum of definite integrals of which, however, all but the term containing $f(r\alpha)$ as a factor will vanish. This particular term reduces to

[1] See Goursat: *Mathematical Analysis* (English Translation, New York), page 364.

$$a\int_{-\pi/a}^{+\pi/a} f(ra)d\omega \qquad \text{or} \qquad 2\pi f(ra)$$

Hence we have

$$f(ra) = \frac{1}{2\pi}\int_{-\pi/a}^{+\pi/a} F(\omega)\, e^{-ra\omega i} d\omega \tag{4a}$$

By letting a converge toward zero and by the substitution $ra = x$ this equation reduces to

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} I(\omega) e^{-x\omega i} d\omega \tag{4b}$$

We then have, if we introduce a new function $\varphi(\omega)$ defined by the simple relation:

$$\sqrt{2\pi}\ \varphi(\omega) = \lim_{a=0} F(\omega), \text{ or}$$

$$\varphi(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x) e^{x\omega i} dx \tag{5a}$$

$$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \varphi(\omega) e^{-x\omega i} d\omega \tag{5b}$$

Charlier has suggested the name *conjugated Fourier function* of $f(x)$ for the expression $\varphi(\omega)$.

The equations (5a) and (5b) are known *as integral equations of the first kind.* The expression $e^{x\omega i}$(or $e^{-x\omega i}$) is known as the *nucleus* of the equation. If in (5b) we know the value of $\varphi(\omega)$ we are able to determine $f(x)$. Inversely, if we know $f(x)$ we may find $\varphi(\omega)$ from (5a).

106. Frequency Function as the Solution of an Integral Equation.—We are now in a position to make use of the semi-invariants of Thiele, which hitherto in our discussion have appeared as a rather disconnected and alien member. On page 191 we saw that the semi-invariants could be expressed by the relation

$$e^{\frac{\lambda_1}{\underline{|1}}\omega + \frac{\lambda_2}{\underline{|2}}\omega^2 + \frac{\lambda_3}{\underline{|3}}\omega^3 + \cdots} = \Sigma\, e^{o_i \omega}$$

when $o_i (i = 1, 2, 3, \ldots)$ denotes the individual observations.

The definition of the *semi-invariants* does not necessitate that all the o's must be different. If some of the o's are exactly alike it is self-evident that the term $e^{o_i\omega}$ must be repeated as often as o_i occurs among all of the observations. If therefore $N\varphi(o_i)$ denotes the absolute frequency of o_i where $\varphi(o_i)$ is the relative frequency function, then the definition of the semi-invariants may be written as:—

$$\sum \varphi(o_i)e^{\frac{\lambda_1}{\lfloor 1}\omega+\frac{\lambda_2}{\lfloor 2}\omega^2+\frac{\lambda_3}{\lfloor 3}\omega^3+\dots} = \sum \varphi(o_i)e^{o_i\omega}$$

For continuous variates, x, the above sums are transformed into definite integrals of the form

$$e^{\frac{\lambda_1}{\lfloor 1}\omega+\frac{\lambda_2}{\lfloor 2}\omega^2+\frac{\lambda_3}{\lfloor 3}\omega^3+\dots}\int_{-\infty}^{+\infty}\varphi(x)dx = \int_{-\infty}^{+\infty}\varphi(x)e^{x\omega}dx$$

Let us now substitute the quantity $\sqrt{-1}\,\omega$, or $i\omega$, for ω in the above identity. We then have:—

$$e^{\frac{\lambda_1}{\lfloor 1}i\omega+\frac{\lambda_2}{\lfloor 2}i^2\omega^2+\frac{\lambda_3}{\lfloor 3}i^3\omega^3+\dots}\int_{-\infty}^{+\infty}\varphi(x)dx = \int_{-\infty}^{+\infty}\varphi(x)e^{ix\omega}dx$$

under the supposition that this transformation holds in the complex region in which the function is defined.

In this equation the definite integrals are of special importance. The factor $\int_{-\infty}^{+\infty}\varphi(x)dx$ is, of course, equal to unity according to the simple considerations set forth on page 189. The integral on the right hand side of the equation is, however, apart from the constant factor $\sqrt{2\pi}$ nothing more than the $\varphi(\omega)$ function in the conjugate Fourier function if we let $\varphi(x) = f(x)$, and

$$e^{\frac{\lambda_1}{\lfloor 1}i\omega+\frac{\lambda_2}{\lfloor 2}i^2\omega^2+\frac{\lambda_3}{\lfloor 3}i^3\omega^3+\dots} = \sqrt{2\pi}\varphi(\omega)$$

According to (5b) we may, therefore write $f(x)$ or $\varphi(x)$ as

$$\varphi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{\frac{\lambda_1}{\lfloor 1}i\omega+\frac{\lambda_2}{\lfloor 2}i^2\omega^2+\frac{\lambda_3}{\lfloor 3}i^3\omega^3+\dots}e^{-x\omega i}d\omega$$

as the most general form of the frequency function $\varphi(x)$ *expressed by means of semi-invariants.* (See Appendix.)

107. The Normal or Laplacean Probability Function.—The exactness with which $\varphi(x)$ is reproduced depends, of course, upon the number of λ's we decide to consider in the above formula. As a first approximation we may omit all λ's above the order 2 or all terms in the exponent with indices higher than 2. Bearing in mind that $i^2 = -1$ we therefore have as a first approximation

$$\varphi_o(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{i\omega(\lambda_1 - x) - \frac{\lambda_2}{\underline{|2}}\omega^2} d\omega.$$

This definite integral was first evaluated by Laplace by means of the following elegant analysis. Using the well known Eulerean relation for complex quantities the above integral may be written as

$$\int_{-\infty}^{+\infty} e^{-\frac{\lambda_2\omega^2}{2}} \cos[(\lambda_1 - x)\omega]\,d\omega + i\int_{-\infty}^{+\infty} e^{-\frac{\lambda_2}{2}\omega^2} \sin[(\lambda_1 - x)\omega]\,d\omega$$

The imaginary number vanishes because the factor $e^{\frac{-\lambda_2}{2}\omega^2}$ is an even function and *sin* $[(\lambda_1 - x)\omega]$ an uneven function, and the area from $-\infty$ to 0 will therefore equal the area from 0 to $+\infty$, but be opposite in sign, which reduces the total area from $-\infty$ to $+\infty$ or the integral in question to zero.

In regard to the first term, similar conditions hold except that $\cos[(\lambda_1 - x)\omega]$ is an even function and the integral may hence be written as

$$I = 2\int_0^{+\infty} e^{-\frac{\lambda_2}{2}\omega^2} \cos(r\omega)d\omega \qquad \text{where } r = \lambda_1 - x$$

Regarding the parameter r as a variable and differentiating I in respect to this variable we have

$$\frac{dI}{dr} = \frac{2}{\lambda_2}\int_0^{+\infty} (-\lambda_2\omega e^{-\frac{\lambda_2}{2}\omega^2}) \sin(r\omega)\,d\omega.$$

From this we have by partial integration:—

$$\frac{dI}{dr} = \frac{2}{\lambda_2}\left[e^{-\frac{\lambda_2}{2}\omega^2} \sin(r\omega)d\omega\right]_0^{+\infty} - \frac{2r}{\lambda_2}\int_0^{\infty} e^{-\frac{\lambda_2}{2}\omega^2}\cos(r\omega)d\omega$$

$$= 0 - \frac{rI}{\lambda_2} \quad \text{or} \quad \frac{1}{I}\frac{dI}{dr} = -\frac{r}{\lambda_2}$$

From which we find

$$\log I = -\frac{r^2}{2\lambda_2} + \log A$$

where $\log A$ is a constant. Hence we have:—

$$I = A\, e^{\frac{-r^2}{2\lambda_2}}$$

In order to determine A we let $r = 0$ and we have

$$I_o = A = 2\int_0^\infty e^{\frac{-\lambda_2}{2}\omega^2} d\omega = 2\sqrt{\frac{\pi}{2\lambda_2}} = \sqrt{\frac{2\pi}{\lambda_2}}$$

This finally gives the expression for $\varphi_o(x)$ in the following form:

$$\varphi_o(x) = \frac{1}{\sqrt{2\pi\lambda_2}}\, e^{-\frac{(\lambda_1 - x)^2}{2\lambda_2}}$$

as a preliminary approximation for the frequency curve $\varphi(x)$.

The first mathematical deduction of this approximate expression for a frequency curve is found in the monumental work by Laplace on Probabilities, and the function $\varphi_o(x)$ entering in the expression $\varphi_o(x)dx$, which gives the probability that the variate will fall between $x - \frac{1}{2}dx$ and $x + \frac{1}{2}dx$, is therefore known as the Laplacean probability function or sometimes as the Normal Frequency Curve of Laplace. The same curve was, as we have mentioned, also deduced independently by Gauss in connection with his studies on the distribution of accidental errors in precision measurements.

Laplace's probability function, $\varphi_o(x)$, possesses some remarkable properties which it might be well worth while to consider. Introducing a slightly different system of notation by writing $\lambda_1 = M$ and $\sqrt{\lambda_2} = \sigma$, $\varphi_o(x)$ reduces to the following form:

$$\varphi_o(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x - M)^2 : 2\sigma^2}$$

which is the form introduced by Pearson.

The frequency curve, $\varphi_o(x)$, is here expressed in reference to a Cartesian coördinate system with origin at the zero point of the natural number system and whose unit of measurement is also equivalent to the natural number unit. It is, however, not necessary to use this system in preference to any other system. In fact, we may choose arbitrarily any other origin and any other unit

standard without altering the properties of the curve. Suppose, therefore, that we take M as the origin and σ as the unit of the system. The frequency function then reduces to

$$\varphi_o(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2:2}$$

where $z = (x - M):\sigma$

Since the integral of $\varphi_o(z)$ from $-\infty$ to $+\infty$ equals unity the following equation must necessarily hold.

$$\int_{-\infty}^{+\infty} e^{-z^2:2} dz = \sqrt{2\pi}$$

This latter result may, however, be deduced independently of the fact that $\varphi_o(z)$ happens to be a probability function. The above definite integral is a form well known from the calculus and equals $\sqrt{2\pi}$. It serves therefore as an independent check of our calculations.

108. Hermite's Polynomials.—The Laplacean Probability Curve possesses, however, some other remarkable properties which are of great use in expanding a function in a series. Starting with $\varphi_o(z)$ we may by repeated differentiation obtain its various derivatives. Denoting such derivatives by $\varphi_1(z)$, $\varphi_2(z)$, $\varphi_3(z)$ . . . respectively we have the following relations.[1]

$$\begin{aligned}
\varphi_0(z) &= e^{-z^2:2} \\
\varphi_1(z) &= -\, z\varphi_0(z) \\
\varphi_2(z) &= (z^2 - 1)\varphi_0(z) \\
\varphi_3(z) &= -\,(z^3 - 3z)\varphi_0(z) \\
\varphi_4(z) &= (z^4 - 6z^2 + 3)\varphi_0(z) \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot
\end{aligned}$$

and in general for the *nth* derivative:—

$$\varphi_n(z) = (-1)^n \Bigg[z^n - \frac{n(n-1)}{2} z^{n-2} + \frac{n(n-1)(n-2)(n-3)z^{n-4}}{2 \cdot 4} - \frac{n(n-1)(n-2)(n-3)(n-4)(n-5)z^{n-6}}{2 \cdot 4 \cdot 6} + \cdot\cdot \Bigg] \varphi_0(z)$$

[1] In the following computations we have omitted temporarily the constant factor $1 : \sqrt{2\pi}$ of $\varphi_o(z)$ and its derivatives.

It can be readily seen that the derivatives of $\varphi_o(z)$ are represented throughout as products of polynomials of z and the function $\varphi_o(z)$ itself. The various polynomials

$$\begin{aligned} H_0(z) &= 1 \\ H_1(z) &= z \\ H_2(z) &= z^2 - 1 \\ H_3(z) &= (z^3 - 3z) \\ H_4(z) &= (z^4 - 6z^2 + 3) \end{aligned}$$

and so forth are generally known as Hermite's polynomials from the name of the French mathematician, Hermite, who first introduced these polynomials in mathematical analysis.

The following relations can be shown to exist

$$H_{n+1}(z) + zH_n(z) + nH_{n-1}(z) = 0$$

and

$$\frac{d^2H_n(z)}{dz^2} - \frac{zdH_n(z)}{dz} + nH_n(z) = 0$$

from which we successively may compute the various $H(z)$.

A numerical 10 decimal place tabulation of the first six Hermite polynomials for values of z up to 4 and progressing by intervals of 0.01 is given by Jörgensen in his aforementioned "Frekvensflader og Korrelation."

109. Orthogonal Functions.—There exist now some very important relations between the Hermite polynomials and the derivatives of $\varphi_o(z)$, or between $H_n(z)$ and $\varphi_n(z)$.

Consider for the moment the two following series of functions

$$\begin{aligned} &\varphi_0(z),\ \varphi_1(z),\ \varphi_2(z),\ \varphi_3(z)\ \varphi_4(z), \dots \\ &H_0(z),\ H_1(z),\ H_2(z),\ H_3(z)\ H_4(z), \dots \end{aligned}$$

where $\varphi_n(z) = (-1)^n H_n(z)\varphi_o(z)$ and where $\lim \varphi_n(z) = 0$ for $z = \pm \infty$

We shall now prove that the two series $\varphi_n(z)$ and $H_n(z)$ form a *biorthogonal* system in the interval $-\infty$ to $+\infty$, that is to say that they are

(1) real and continuous in the whole plane
(2) no one of them is identically zero in the plane
(3) every pair of them, $\varphi_n(z)$ and $H_m(z)$ satisfy the relation

$$\int_{-\infty}^{+\infty} \varphi_n(z)H_m(z)dz = 0 \qquad (n \gtrless m)$$

We have the self-evident relation

$$\int_{-\infty}^{+\infty} H_m(z)\varphi_n(z)dz = \int_{-\infty}^{+\infty} H_m(z)H_n(z)\varphi_o(z)dz = \int_{-\infty}^{+\infty} H_n(z)\varphi_m(z)dz$$

Since this relation holds for all values of m and n it is only necessary to prove the proposition for $n > m$. For if it holds for $n > m$ it will according to the above relation also hold for $n < m$.

By partial integration we have:—

$$\int_{-\infty}^{+\infty} H_m(z)\varphi_n(z)dz = \Big[H_m(z)\varphi_{n-1}(z)\Big]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} H_m'(z)\varphi_{n-1}(z)dz$$

when $H_m'(z)$ is the first derivative of $H_m(z)$.

The first member on the right reduces to 0 since $\varphi_{n-1}(z) = 0$ for $z = \pm\infty$ and because H_m is of a lower order than φ_n. We have therefore:—

$$\int_{-\infty}^{+\infty} H_m(z)\varphi_n(z)dz = -\int_{-\infty}^{+\infty} H_m'(z)\varphi_{n-1}(z)dz$$

$$\int_{-\infty}^{+\infty} H_m'(z)\varphi_{n-1}(z)dz = -\int_{-\infty}^{+\infty} H_m''(z)\varphi_{n-2}(z)dz$$

$$\int_{-\infty}^{+\infty} H_m''(z)\varphi_{n-2}(z)dz = -\int_{-\infty}^{+\infty} H_m'''(z)\varphi_{n-3}(z)dz$$

Continuing this process we obtain finally an expression of the form

$$\int_{-\infty}^{+\infty} H_m(z)\varphi_n(z)dz = (-1)^{m+1}\int_{-\infty}^{+\infty} H_m^{(m+1)}\varphi_{n-m-1}(z)dz$$

where $H_m^{(m+1)}(z)$ is the $m+1$ derivative of $H_m(z)$ and $n - m - 1 \geqq 0$. Since $H_m(z)$ is a polynomial in the m^{th} degree its $m+1st$ derivative is zero, and we have finally that

$$\int_{-\infty}^{+\infty} H_m(z)\varphi_n(z)dz = 0$$

for all values of m and n where $n \gtrless m$.

For $m = n$ we proceed in exactly the same manner, but stop at the $m^{\underline{th}}$ integration. We have, therefore, by replacing m by n in the above partial integrations

$$\int_{-\infty}^{+\infty} H_n(z)\varphi_n(z)dz = (-1)^n \int_{-\infty}^{+\infty} H_n^{(n)}(z)\varphi_{n-n}(z)dz =$$

$$(-1)^n \int_{-\infty}^{+\infty} H_n^{(n)}(z)\varphi_o(z)dz$$

The $n^{\underline{th}}$ derivative of $\text{H}_n(z)$ is however nothing but a constant and equal to $\underline{|n}$. Hence we have finally

$$\int_{-\infty}^{+\infty} H_n(z)\varphi_n(z)dz = (-1)^n\underline{|n} \int_{-\infty}^{\infty +} e^{-z^2:2}dz = (-1)^n \underline{|n}\sqrt{2\pi}$$

The above analysis thus proves that the functions $H_m(z)$ and $\varphi_n(z)$ are biorthogonal to each other for all values of n different from m throughout the whole plane.

110. The Frequency Function expressed as a Series.—We can now make use of these relations between the infinite set of biorthogonal functions $H_m(z)$ and $\varphi_n(z)$ in solving the problem of expanding an arbitrary function $\varphi(z)$ in a series of the form

$$\varphi(z) = c_o\varphi_o(z) + c_1\varphi_1(z) + c_2\varphi_2(z) + \ldots,$$

the series to hold in the interval from $-\infty$ to $+\infty$.

If we know that $\varphi(z)$ can be developed into a series of this form, which after multiplication by any continuous function can be integrated term for term, then we are able to give a formal determination of the coefficients c.

This formal determination of any one of the c's, say c_i, consists in multiplying the above series by $H_i(z)$ and integrating each term from $-\infty$ to $+\infty$. All the terms except the one containing the product $H_i(z)\varphi_i(z)$ vanish and we have for c_i.

$$c_i = \frac{\int_{-\infty}^{+\infty} \varphi(z)H_i(z)dz}{\int_{-\infty}^{+\infty} \varphi_i(z)H_i(z)dz} = \frac{\int_{-\infty}^{+\infty} \varphi(z)H_i(z)dz}{(-1)^i\underline{|i}\sqrt{2\pi}}$$

It will be noted that this purely formal calculation of the co-

efficients c_i is very similar to the determination of the constants in a Fourier Series, where as a matter of fact the system of functions

$$\cos z, \cos 2z, \cos 3z \ldots$$
$$\sin z, \sin 2z, \sin 3z \ldots$$

is biorthogonal in the interval $0 \leqq z \leqq 1$.

But the reader must not forget that the above representation is only a formal one, and we do not know if it is valid. To prove its validity we must first show that the series is convergent and secondly that it actually represents $\varphi(z)$ for all values of z.

This is by no means a simple task and it cannot be done by elementary methods. A Russian mathematician, Vera Myller-Lebedeff, has, however, given an elegant solution by means of some well-known theorems from the Fredholm integral equations. She has among other things proved the following criterion:—

"Every function $\varphi(z)$ which together with its first two derivatives is finite and continuous in the interval from $-\infty$ to $+\infty$ and which vanishes together with its derivatives for $z = \pm\infty$ can be developed into an infinite series of the form:—

$$\varphi(z) = \sum c_i e^{-z^2:2} H_i(z)$$

where $H_i(z)$ is the Hermite polynomial of order i."

111. Derivation of Gram's Series.—It is, however, not our intention to follow up this treatment which is outside the scope of an elementary treatise like this and shall in its place give an approximate representation of the frequency function, $\varphi(z)$, by a method, which in many respects is similar to that introduced by the Danish actuary Gram in his epoch-making work "Udviklingsraekker," which contains the first known systematic development of a skew frequency function. Gram's problem in a somewhat modified form may briefly be stated as follows:—*Being given an arbitrary relative frequency function, $\varphi(z)$, continuous and finite in the interval $-\infty$ to $+\infty$ (and which vanishes for $z = \pm\infty$) to determine the constant coefficients c_0, c_1, c_2, c_3 . . . in such a way that the series*

$$\frac{c_0\varphi_0(z)}{\sqrt{\varphi_0(z)}} + \frac{c_1\varphi_1(z)}{\sqrt{\varphi_0(z)}} + \frac{c_2\varphi_2(z)}{\sqrt{\varphi_0(z)}} + \ldots \frac{c_n\varphi_n(z)}{\sqrt{\varphi_0(z)}}$$
$$= \frac{1}{\sqrt{\varphi_0(z)}} \sum c_i\varphi_i(z)$$

gives the best approximation to the quantity $\varphi(z) : \sqrt{\varphi_o(z)}$ *in the sense of the method of least squares.* That is to say we wish to determine the constants c in such a manner that the sum of the squares of the differences between the function and the approximate series becomes a minimum. This means that the expression

$$I = \int_{-\infty}^{+\infty} \left[\frac{\varphi(z)}{\sqrt{\varphi_o(z)}} - \sum \frac{c_i\varphi_i(z)}{\sqrt{\varphi_o(z)}}\right]^2 dz$$

must be a minimum.

On the basis of this condition we have

$$\frac{\varphi(z)}{\sqrt{\varphi_o(z)}} \lesseqgtr \frac{1}{\sqrt{\varphi_o(z)}} \sum c_i\varphi_i(z) = \sqrt{\varphi_o(z)} \sum c_i H_i(z) = U(z)$$

where the unknown coefficients c must be so determined that

$$I = \int_{-\infty}^{+\infty} \left[\frac{\varphi(z)}{\sqrt{\varphi_o(z)}} - U(z)\right]^2 dz \text{ is a minimum}$$

Taking the partial derivatives with respect to c_i we have

$$\frac{\delta I}{\delta c_i} = -\frac{2\delta}{\delta c_i}\int_{-\infty}^{+\infty} \frac{\varphi(z)}{\sqrt{\varphi_o(z)}} U(z)dz + \frac{\delta}{\delta c_i}\int_{-\infty}^{+\infty} \left[U(z)\right]^2 dz$$

Now since

$$\int_{-\infty}^{+\infty} \left[U(z)\right]^2 dz = \int_{-\infty}^{+\infty} \left\{ c_o^2\left[H_o(z)\right]^2 + c_1^2\left[H_1(z)\right]^2 + \ldots c_n^2\left[H_n(z)\right]^2 \right\} \varphi_o(z)dz$$

we get

$$\frac{\delta I}{\delta c_i} = -2\int_{-\infty}^{+\infty} \frac{\varphi(z)}{\sqrt{\varphi_o(z)}} H_i(z)\sqrt{\varphi_o(z)}dz + 2c_i\int_{-\infty}^{+\infty} \left[H_i(z)\right]^2 \varphi_o(z)dz$$

where the latter integral equals $\int_{-\infty}^{+\infty} \varphi_i(z)H_i(z)dz = (-1)^i \underline{|i}\sqrt{2\pi}$

Equating to zero and solving for c_i we finally obtain the following value for c_i

$$c_i = \frac{(-1)^i}{\underline{|i}\sqrt{2\pi}} \int_{-\infty}^{+\infty} \varphi(z)H_i(z)dz \qquad \text{for } i = 1, 2, 3, \ldots$$

This solution is gotten by the introduction of $\sqrt{\varphi_o(z)}$ which serves to make all terms of the form $c_i\varphi_i(z) : \sqrt{\varphi_o(z)}$ equal to $\sqrt{\varphi_o(z)}c_iH_i(z)$ ($i = 1, 2, 3, \ldots n$) orthogonal to each other in the interval $-\infty$ to $+\infty$.

In all the above expansions of a frequency series we have used the expression $\varphi_o(z) = e^{-z^2/2}$ as the generating function (see footnote on page 199), while as a matter of fact the true value of $\varphi_o(z)$ is given by the equation $\varphi_o(z) = e^{-z^2/2} : \sqrt{2\pi}$

The definite integral on page (202)

$$(-1)^i \int_{-\infty}^{+\infty} H_i(z)\varphi_i(z)dz = \underline{|i} \int_{-\infty}^{+\infty} e^{-z^2:2}dz = \underline{|i}\sqrt{2\pi}$$

will therefore have to be divided by $\sqrt{2\pi}$, and the value of the general coefficient c_i will henceforth be reduced to

$$c_i = \frac{\int_{-\infty}^{+\infty} \varphi(z)H_i(z)dz}{(-1)^i\underline{|i}}$$

where $H_i(z)$ is the Hermite polynomial of order i defined by the relation

$$H_i(z) = z^i - \frac{i(i-1)z^{i-2}}{2} + \frac{i(i-1)(i-2)(i-3)}{2\cdot 4}z^{i-4}$$
$$\frac{-\,i(i-1)(i-2)(i-3)(i-4)(i-5)z^{i-6}}{2\cdot 4\cdot 6} + \ldots$$

On this basis we obtain the following values for the first four coefficients:—

$$c_0 = \int_{-\infty}^{+\infty} \varphi(z)dz = 1$$

$$c_1 = (-1)^1 \int_{-\infty}^{+\infty} \varphi(z)zdz : \underline{|1}$$

$$c_2 = (-1)^2 \int_{-\infty}^{+\infty} (z^2 - 1)\varphi(z)dz : \underline{|2}$$

$$c_3 = (-1)^3 \int_{-\infty}^{+\infty} (z^3 - 3z)\varphi(z)dz : \underline{|3}$$

$$c_4 = (-1)^4 \int_{-\infty}^{+\infty} (z^4 - 6z^2 + 3z)\varphi(z)dz : \underline{|4}$$

112. Absolute Frequencies.—While the above development of an arbitrary frequency distribution has reference to $\varphi(z)$, or the relative frequency function, it is, however, equally well adapted to the representation of absolute frequencies as expressed by the function, $F(z)$. If N is the total number of individual observations, or in other words the area of the frequency curve, we evidently have

$$F(z) = N\varphi(z) \text{ or} \int_{-\infty}^{+\infty} F(z)dz = N\int_{-\infty}^{+\infty} \varphi(z)dz = N.$$

Since N is a constant quantity we may, therefore, write the expansion of $F(z)$ as follows:

$$F(z) = N\left[c_o\varphi_o(z) + c_1\varphi_1(z) + c_2\varphi_2(z) + \dots\right]$$

$$= N\sum c_i H_i(z)e^{-z^2:2}$$

where the coefficients c_i have the value

$$c_i = \frac{(-1)^i}{N\underline{|i}} \int_{-\infty}^{+\infty} F(z)H_i(z)dz, \text{ for } i = 1, 2, 3, \dots$$

and where

$$N = \int_{-\infty}^{+\infty} F(z)dz.$$

Since all the Hermite functions are polynomials in z it can be readily seen that the coefficients c may be expressed as functions of the power sums or of the previously mentioned symmetrical functions s, where

$$s_r = \int_{-\infty}^{+\infty} z^r F(z)dz$$

These particular integrals originally introduced by Thiele in the development of the semi-invariants have been called by Pearson the "*moments*" of the frequency function, $F(z)$, and s_r is called the *rth* moment of the variate z with respect to an arbitrary origin.

It can be readily seen that the moment of order zero or s_o is

$$s_o = \int_{-\infty}^{+\infty} z^o F(z)dz = N \int_{-\infty}^{+\infty} \varphi(z)dz = N$$

Hence we have for the first coefficient c_o

$$c_o = \int_{-\infty}^{+\infty} F(z)dz : \int_{-\infty}^{+\infty} F(z)dz = 1$$

We are, however, in a position to further simplify the expression for $F(z)$.

As already mentioned we are at liberty to choose arbitrarily both the origin and the unit of the Cartesian coördinate system for the frequency curve without changing the properties of this curve. Now by making a proper choice of this Cartesian system of reference we can make the coefficients c_1 and c_2 vanish. In order to obtain this object the origin of the system must be so chosen that

$$c_1 = \frac{-1}{\underline{|1}} \int_{-\infty}^{+\infty} zF(z)dz : \int_{-\infty}^{+\infty} F(z)dz = 0$$

This means that the semi-invariant $s_1 : s_0 = \lambda_1$ must vanish. It can be readily seen that the above expression for λ_1 is nothing more than the usual form for the mean value of a series of variates.

Moreover, we know that the algebraic sum (or in the case of continuous variates, the integral) of the variates around the mean value is always equal to zero. Hence by writing for z the expression $(z - M)$ when M equals the mean value or λ_1, we can always make c_1 vanish.

To attain our second object of making c_2 vanish we must choose the unit of the coördinate system in such a way that the expression

$$c_1 = \frac{(-1)^2}{\underline{|2}} \int_{-\infty}^{+\infty} F(z)H_2(z)dz : \int_{-\infty}^{+\infty} F(z)dz = 0$$

which implies that

$$\left[\int_{-\infty}^{+\infty} F(z)z^2dz - \int_{-\infty}^{+\infty} F(z)dz\right] : \int_{-\infty}^{+\infty} F(z)dz = 0$$

or that $s_2 : s_1 - 1 = 0$, or when expressed in terms of the semi-invariants that

$$\lambda_2 = (s_2s_0 - s_1^2) : s_0^2 = 1.$$

But by choosing the mean as the origin of the system the term $s_1 : s_0$ is equal to 0 and we have therefore $\lambda_2 = \sigma^2 = s_2 : s_0 = 1.$ Hence, by selecting as the unit of our coördinate system $\sqrt{\lambda_2}$ or σ, where σ is technically known as the dispersion or standard deviation of the series of variates, we can make the second coefficient c_2 vanish.

In respect to the coefficients c_3 and c_4 we have now

$$c_3 = \frac{(-1)^3}{\underline{|3}}\left[\int_{-\infty}^{+\infty} z^3F(z)dz - 3 \int_{-\infty}^{+\infty} zF(z)dz\right] : \int_{-\infty}^{+\infty} F(z)dz$$

which reduces to $-\frac{1}{\underline{|3}}\frac{s_3}{s_0}$, while

$$c_4 = \frac{(-1)^4}{\underline{|4}}\left[\int_{-\infty}^{+\infty} z^4F(z)dz - 6\int_{-\infty}^{+\infty} z^2F(z)dz + 3\int_{-\infty}^{+\infty} F(z)dz\right] : \int_{-\infty}^{+\infty} F(z)dz,$$

which reduces to

$$\frac{1}{\underline{|4}}\left[\frac{s_4}{s_0} - \frac{6s_2}{s_0} + \frac{3s_0}{s_0}\right] = \frac{1}{\underline{|4}}\left[\frac{s_4}{s_0} - 3\right]$$

While the coefficients of higher order may be determined with equal ease it will in general be found that the majority of mod-

erately skew frequency distributions can be expressed by means of the first 4 parameters or coefficients.

113. Coefficients expressed by Semi-Invariants.—We shall now show how the same results for the values of the coefficients may be obtained from the definition of the semi-invariants. Since we have proven that a frequency function, $F(z)$, may be expressed by the series

$$F(z) = \Sigma c_i \varphi_i(z)$$

we may from the definition of the semi-invariants write down the following identity:—

$$s_o e^{\frac{\lambda_1 \omega}{\underline{|1}} + \frac{\lambda_2 \omega^2}{\underline{|2}} + \cdots} = N \int_{-\infty}^{+\infty} e^{z\omega} (c_o \varphi_o(z) + c_1 \varphi_1(z) + c_2 \varphi_2(z) + \ldots dz$$

where N is the area of the frequency curve.

The general term on the right hand side of the equation will be of the form

$$c_r \int_{-\infty}^{+\infty} e^{z\omega} \varphi_r(z) dz$$

where the integral may be evaluated by partial integration as follows:—

$$\int_{-+}^{+\infty} e^{z\omega} \varphi_r(z) dz = e^{z\omega} \varphi_{r-1}(z) \Big]_{-\infty}^{+\infty} - \omega \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-1}(z) dz,$$

and where the first term on the right vanishes leaving

$$\int_{-\infty}^{+\infty} e^{z\omega} \varphi_r(z) dz = (-\omega)^1 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-1}(z) dz$$

Continuing in the same manner we obtain by successive integrations

$$(-\omega)^1 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-1}(z) dz = (-\omega)^2 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-2}(z) dz$$

$$(-\omega)^2 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-2}(z) dz = (-\omega)^3 \int_{-\infty}^{+\infty} e^{z\omega} \varphi_{r-3}(z) dz$$

. . . .

. . . .

from which we finally obtain the relation

$$\int_{-\infty}^{+\infty} e^{z\omega}\varphi_r(z)dz = (-\omega)^r \int_{-\infty}^{+\infty} e\,\varphi_o(z)dz = \frac{(-\omega)^r}{\sqrt{\pi 2}} \int_{-\infty}^{+\infty} e^{z\omega - \frac{z^2}{2}} dz$$

This latter integral may be written as

$$\frac{(-\omega)^r}{\sqrt{\pi 2}} e^{\frac{\omega^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z-\omega)^2} dz = \frac{(-\omega)^r}{\sqrt{2\pi}} e^{\frac{\omega^2}{2}} \sqrt{2\pi} = (-\omega)^r e^{\frac{\omega^2}{2}}$$

Consequently the relation between the semi-invariants and the frequency function may be written as follows:—

$$s_o e^{\frac{\lambda_1\omega}{|1} + \frac{\lambda_2\omega^2}{|2} + \ldots} = N\left[c_o c_1\omega + c_2\omega^2 - c_3\omega^3 + \ldots\right] e^{\frac{\omega^2}{2}}, \text{ or}$$

$$s_o e^{\frac{\lambda_1\omega}{|1} + \frac{\omega^2}{|2}(\lambda_2 - 1) + \ldots} = N\left[(c_o - c_1\omega + c_2\omega^2 - c_3\omega^3 + \ldots\right].$$

By successive differentiation with respect to ω and by equating the coefficients of equal powers of ω we get in a manner similar to that shown on page 192 the following results:

$$c_0 = s_0 : N = s_0 : s_0 = 1$$

$$c_1 = -\lambda_1$$

$$c_2 = \frac{1}{|2}\left[(\lambda_2 - 1) + \lambda_1^2\right]$$

$$c_3 = \frac{-1}{|3}\left[\lambda_3 + 3(\lambda_2 - 1)\lambda_1 = \lambda_1^3\right]$$

$$c_4 = \frac{1}{|4}\left[\lambda_4 + 4\lambda_3\lambda_1 + 3(\lambda_2 - 1)^2 + 6(\lambda_2 - 1)\lambda_1^2 + \lambda_1^4\right]$$

If we now again choose the origin at λ_1 or let $\lambda_1 = 0$ and choose $\sqrt{\lambda_2} = 1$ as the unit of our coördinate system we have:—

$$c_0 = 1, c_1 = 0, c_2 = 0, c_3 = \frac{-1}{|3}\lambda_3, c_4 = \frac{1}{|4}\lambda_4$$

114. Change of Origin and Unit.—The theoretical development of the above formulæ explicitly assumes that the variate, z, is measured in terms of the dispersion or $\sqrt{\lambda_2(z)}$ and with $\lambda_1(z)$

as the origin of the coördinate system. In practice the observations or statistical data are, however, invariably expressed with reference to an arbitrarily chosen origin (in the majority of cases the natural zero of the number scale) and expressed in terms of standard units, such as centimeters, grams, years, integral numbers, etc.

Let us denote the general variate in such arbitrarily selected systems of reference by x. Our problem then consists in transforming the various semi-invariants, $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$, . . . to the system of reference with $\lambda_1(z)$ as its origin and $\sqrt{\lambda_2(z)}$ as its unit. Such a transformation may always be brought about by means of the linear substitution

$$z = ax + b$$

which in a purely geometrical sense implies both a change of origin and unit. On page 193 we proved the following general properties of the semi-invariants

$$\begin{aligned} \lambda_1(z) &= \lambda_1(ax + b) = a\lambda_1(x) + b \\ \lambda_r(z) &= \lambda_r(ax + b) = a^r\lambda_r(x) \end{aligned}$$

Let us now write $\lambda_1(x) = M$ and $\lambda_2(x) = \sigma^2$, we then have the following relations:—

$$\begin{aligned} \lambda_1(z) &= aM + b \\ \lambda_2(z) &= a^2\sigma^2 \end{aligned}$$

Since the coördinate system of reference must be chosen in such a manner that $\lambda_1(z) = 0$ and $\sqrt{\lambda_2(z)} = 1$ we have

$$\begin{aligned} aM + b &= 0 \\ a\sigma \quad\;\; &= 1 \end{aligned}$$

from which we obtain $a = \dfrac{1}{\sigma}$ and $b = \dfrac{-M}{\sigma}$, which brings z on the form $z = (x - M) : \sigma$, while $\varphi_o(z)$ becomes

$$\varphi_o(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-M)^2 : 2\sigma^2}$$

Moreover, we have $\lambda_r(z) = \lambda_r(x) : \sigma^r$ for all values of r greater than 2. We are now able to epitomize the computations of the semi-invariants under the following simple rules:

(1) Compute $\lambda_1(x)$ in respect to an arbitrary origin. The numerical value of this parameter with opposite sign is the origin of the frequency curve.

(2) Compute $\lambda_2(x)$ for all values of r equal or greater than 2. The numerical values of those parameters divided with $(\sqrt{\lambda_2(x)})^r$ or σ^r, for $r=2, 3, 4, \ldots$ are the semi-invariants of the frequency curve.

Remarks on Nomenclature and Tables.—We shall now briefly discuss some of the geometrical properties of the Laplacean probability curve $\varphi_0(z)=e^{-z^2:2}$ and its derivatives, $\varphi_i(z)=H_i(z)\varphi_0(z)$, for $i=3, 4, 5 \ldots$ Writing $\varphi_0(z)$ and its derivatives as:

$$\begin{aligned}
\varphi_0(z) &= e^{-z^2:2}:\sqrt{2\pi}\\
\varphi_1(z) &= -z\varphi_0(z)\\
\varphi_2(z) &= (-1)^2(z^2-1)\varphi_0(z)\\
\varphi_3(z) &= (-1)^3(z^3-3z)\varphi_0(z)\\
\varphi_4(z) &= (-1)^4(z^4-6z^2+3)\varphi_0(z)\\
&\cdots\cdots\cdots\\
&\cdots\cdots\cdots
\end{aligned}$$

we readily notice that both $\varphi_0(z)$ and all its derivatives of even order are even functions of z while all the derivatives of uneven order are uneven functions.

The Laplacean probability function which occurs as a factor in all the expressions is in itself a single valued positive function with a maximum point at $z=0$ and a point of inflection at $z=\pm 1$ and approaches the abscissa axis asymptotically in both positive and negative direction. At $z=0$ we have $\varphi_0(z)=1:\sqrt{2\pi}=0.3989$. At plus or minus one, $\varphi_0(z)$ is less than 0.25 at $z=\pm 2$, $\varphi_0(z)$ is nearly 0.05, at plus or minus 3 about 0.004 and at $z=\pm 4$ only 0.0001.

In regard to the third derivative, $\varphi_3(z)=H_3(z)\varphi_0(z)$, we find that it possesses a maximum or minimum in the neighborhood of $z=+0.7$ and $z=-0.7$ respectively, it crosses the abscissa axis in the neighborhood of the points $z=\pm 1.75$ and approaches the abscissa asymptotically in both positive and negative direction.

The fourth derivative has a major maximum point at $z=0$, it crosses the abscissa axis from positive to negative direction in the neighborhood of $z=\pm 0.75$, attains a minimum at about $z=\pm 1.35$, it crosses again the abscissa (this time from negative to positive direction) in the neighborhood of $z=\pm 2.3$, attains a secondary or minor maximum around $z=\pm 2.86$ and begins then to decline until it ultimately approaches the abscissa axis asymptotically.

These geometrical properties of the Laplacean frequency curve and its derivatives are, however, much more readily vizualized in the accompanying diagram which needs no further explanation. We wish, however, to call the attention of the reader to the wavelike form of the various curves, which is strongly reminiscent of the form of functions encountered

in harmonic analysis or in the expansions in Fourier Series, an analogy which we had occasion to mention in the discussion of the orthogonal properties of the Hermite polynomials and the derivatives of the Laplacean function.

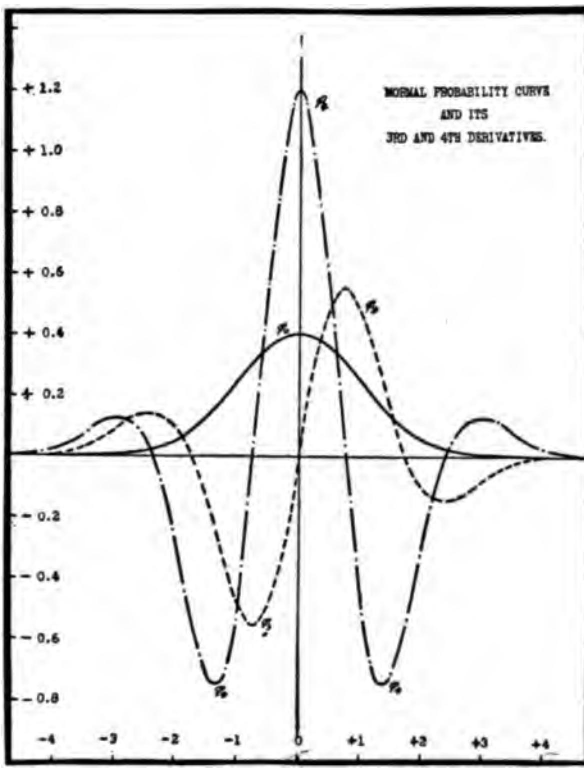


FIGURE 2

In order to facilitate practical numerical calculations it is, however, necessary to have an extensive set of numerical tables for $\varphi_0(z)$ and its derivatives. This fact was already noted by Laplace who more than 25 years prior to the publication of the memoirs by Gauss on the normal error curve advocated the construction of a table of numerical values of the integral.

$$\frac{1}{\sqrt{2\pi}}\int_0^z e^{-z^2:2}dz^*$$

The first set of such tables was constructed by the astronomer, Kramp and modified forms of these tables are found in nearly all treatises on least squares and standard texts on probabilities. The most recent set of

*This fact, as pointed out by Pearson, definitely establishes Laplace's priority of discovery of the probability curve.

tables of this integral are those of Sheppard in (*Tables for Biometricians*, edited by Karl Pearson) where the variate z is expressed in units of σ, or the dispersion. Sheppard has also computed a table of the numerical values of $\varphi_0(z)$.

In order to use the Gram—Charlier expansion in serial form, it is, however, necessary to compute tables for the derivatives—up to the fourth order. A brief table of the first 6 derivatives is already found in Thiele's earlier treatises. Charlier was, however, the first to supply an extensive set of tables to 4 decimal places for values of z up to 4 and progressing by intervals of 0.01 in his *Researches on the Theory of Probabilities* in the *Meddelande* for 1904. The most detailed tables are those of Jörgensen in his *Frekvensflader og Korrelation* which gives the values of $\varphi_0(z)$ and its first 6 derivatives to 7 decimal places for values of z up to 4 and progressing by intervals of 0.01.† The German astronomer Bruhns has in his *Kollektivmasslehre* given a set of tables to 4 decimal places of the values of the definite integrals

$$\int_0^z \varphi_i(z)dz \text{ for } i = 0, 1, 2, 3, 4, 5$$

The Gram-Charlier series gives us the frequency function in the form $F(z) = \sum c_i\varphi_i(z)$ where the various coefficients c_i are expressed as moments or semi-invariants. As we have already pointed out the derivatives of uneven order are uneven functions and the derivatives of even order are even functions. The addition of such terms as $c_3\varphi_3(z)$, $c_5\varphi_5(z)$, . . . tends therefore to produce asymmetry or skewness from the normal form, while addition of the terms $c_4\varphi_4$, $c_6\varphi_6$, . . does not alter the symmetrical form but tends to make it topheavy or flatten it around the neighborhood of the origin or mean value of the variate z. The coefficient c_3:3! (or λ_3: 3!) is technically known as the *skewness*, and c_4:4! (or λ_4:4!) as the *excess* of the curve. No particular names have as yet been proposed for the semi-invariants of higher order.

†I intend to publish a similar set of 5 decimal place tables in the second volume of this treatise.

PART III

PRACTICAL APPLICATIONS OF THE THEORY

Note:—In the following pages the factorial $\underline{|n} = 1 \cdot 2 \cdot 3 \cdots n$ is replaced by the symbol $n!$ and the exponent $\frac{1}{2}n^2$ in the exponential expression $e^{m+\frac{1}{2}n^2}$ must be interpreted as $n^2:2$ and *not* as $1:2n^2$.

CHAPTER XV.

THE NUMERICAL DETERMINATION OF THE PARAMETERS.

115. General Remarks.—The previous investigations on frequency functions have all been more or less of a purely theoretical nature. In the present chapter we now propose to show how the parameters are determined in actual practice from the individual observations or statistical summaries

The determination of these unknown co-efficients or parameters can—as emphasized by Jørgensen in his *Frekvensflader og Korrelation*—be looked at from two points of view. We may either consider the series as infinite in which case the question of determining the co-efficients becomes a problem in the Theory of Functions; or we may decide to consider a finite number of terms in the series and determine the coefficients so that the sum of the squares of the deviations of the resulting function from the observed statistical data becomes a minimum in the sense of the method of least squares. In this case the coefficients and not the moments or semi-invariants are representative of the observations. This latter method is the classical method as used by Gram in his fundamental research on the expansion of frequency functions in series. A brief statement of the essential differences of the two methods may, however, be of advantage to the reader.

The method of moments requires that the areas of the definite integrals of the form $\int_{-\infty}^{+\infty} x^r F(x)dx$ must equal the areas of the observations which are expressed as power sums of the form

$$\sum_{x=-\infty}^{x=\infty} x^r F(x)$$

while the method of least squares requires that

$$\int_{-\infty}^{+\infty} [F(x) - \sum c_i \varphi_i(x)]^2 dx$$

must equal a minimum but does not necessarily impose any restrictions as to the condition of equality of the observed and the computed areas as derived from the mathematical formula.

The problem of determining the parameters in the sense of the method of least squares is therefore essentially a simple problem in maximum and minimum and is not necessarily—as some critics have imagined it to be—invariably interwoven with the law of errors as expressed by the Laplacean probability curve. It is, of course, true that the law of errors can be proved by regarding the principles of the method of least squares as an axiom, and inversely by accepting the law of errors as an axiom, *i. e.* by assuming the deviations from the observations and the functional law or mathematically determined frequency curve as being due to chance or random sampling, we may prove that the sum of the squares of such deviations actually becomes a minimum. This peculiar relation is, however, not necessary or required when we view the determination of the constants as a simple problem in maxima and minima.

116. Remarks on Certain Criticisms.—Many English and American actuaries have of late shown a tendency to ignore the method of least squares and prefer to rely entirely upon the method of moments. Thus Palin Elderton in his otherwise useful and instructive work on *Frequency Curves and Correlation* states that the method is of little practical use, while Mr. D. Caradog Jones in his newly published *First Course in Statistics* claims the method of least squares "which is the traditional way of approaching all such problems, is shown to be impracticable in a large number of cases, either because the resulting equations cannot be solved, or, when they are capable of solution, because the labour involved would be colossal." This objection falls, however, to the ground in the case of the expansion of a frequency function in serial form because the unknown parameters, with the exception of the origin (the mean) and the unit (the dispersion) of the co-ordinate system, all appear as coefficients in true linear equations and hence are eminently adaptable to the treatment by least squares.

The attitude of these writers is probably due to the fact that they work exclusively with the Pearsonian type of frequency curves where the function, $F(z)$, is given as a closed expression rather than as an expansion in serial form. In nearly all Pearson's curve types there appear not more than four constants which in a measure accounts for the often successful application of the method of moments, although several of the examples presented by Mr. Jones in his book can scarcely be said to be recommendative to Pearson's theory. On the other hand, it is a great drawback, not being able to have more than four constants at our disposal. Personally I have encountered a large number of statistical series where the Pearsonian theory fails. This same fact is also noted by Jorgensen who on page 39 of his *Frekvensflader og Korrelation* states that "jeg kender flere Iagttagelsesraekker, hvor Pearson's Teori svigter totalt."

In the purely theoretical development it matters but little whether we use moments or least squares in the expansion of a frequency function in a series; a fact which is readily seen from our previous demonstrations. In the purely practical work we have, however, this fact to consider,

that the method of moments works exclusively with areas expressed as definite integrals, which are often difficult to determine in extremely skew distributions. And it is only by successive approximations that we in this manner reach a plausible result. Moreover, unless the observations are very numerous, it is almost hopeless to compute the moments of higher order than the fourth, because of the very large errors arising from random sampling. Charlier in one of his monographs asserts that it is generally useless to compute moments of higher order than the second when the number of individual observations in the statistical series is less than 1000. Thiele gives the following brief rules:

For the first and second semi-invariants rely exclusively on the observed data.

For semi-invariants of higher order than 6 rely exclusively on theoretical considerations.

For intermediate semi-invariants (between the 2d and 6th) rely partly upon theory and partly upon the observations.

Caradog Jones, on the other hand, lustily ventures forth with moments of the fourth order, based upon 241, and in some instances even as low as 180 individual observations. It is, therefore, no wonder that some of his results exhibit a somewhat poor "fit" with the original data. Another criticism which may be lodged against the method of moments as used by some adherents of the Pearsonian school, rather than by Pearson himself, is that it works with unweighted observations, and the values of the extremities of the frequency curves are given the same weights as the more numerous observations in the immediate neighborhood of the mean.

A second objection, raised among others by Elderton, is that the expansion in serial form sometimes gives rise to negative frequencies at the extreme tail ends of the curve, due of course to the fact that we have used a limited number of terms of the series. From purely practical considerations this objection counts little, because the observations at the extremeties are very few in number. It matters, for instance, but little in ordinary calculations of assurance premiums whether the upper limit of a mortality table is at 90 or at 100, and when Pearson from his curves actually has attempted to put an upper limit to the duration of human life, he has, to borrow an expression from the Danish biologist, Johannsen, begun to handle biology *as* mathematics and not *with* mathematics. In this connection it may also be noted that the Pearson Type I curve gives imaginary values beyond certain limits. When now certain followers of the Pearsonian school have considered this as an advantage and tried to interprete the limits as possbile values of repeated or presumptive observations, it seems that such disciples have stretched their point a bit too far. It is not possible to see why negative results should be less plausible than imaginary results. Every student of ordinary algebra knows that the "imaginary" quantities are just as valid as the so-called "real" quantities, and it is probably the choice of this unhappy and ill-chosen nomenclature which has given rise to the above extravagant claims of some of the followers of Pearson.

Finally some English and American actuaries have objected to the arbitrary choice of the parameters in the Gram or Charlier expansions. Unless I have completely misunderstood Mr. Elderton this is one of his his chief criticisms against Charlier's method. With my best intentions I cannot agree to this and will even go so far as to say that Mr. Elderton's criticism really speaks in favor of the methods put forth by the Scandinavian scholars. As we have repeatedly emphasized in the preceding paragraphs, the arbitrary choice of c_1 and c_2 amounts mathematically to the choice of an arbitrary origin and unit in the Cartesian co-ordinate system to which surely no mathematician will make objections. Neither can objections be raised from the point of view of common sense. We might as well object to the meter as a unit of measure in preference to the yard, or to reckoning the solar time from the Greenwich meridian instead of the meridian of Paris.

The failure of the method of moments to compute with any degree of accuracy moments of higher order in the case of the majority of ordinary observations is probably the reason why some actuaries, especially in America, have maintained that the Gram or Charlier A type of frequency curves is not powerful enough to represent more than moderately skew frequency distributions.

In spite of the incontrovertible fact that the most recent researches in the theory of integral equations have demonstrated beyond doubt that any frequency curve can be developed in convergent series by Hermite polynomials in conjunction with the normal Laplacean frequency curve an American actuary, Mr. Merwyn Davis, has taken the "bull by the horns" so to speak and boldly gone on record with the positive statement that "the Charlier series fails completely in cases of appreciable skewness." With all due respect for this young matador who has so boldly entered the ring to challenge the work of some of the most eminent mathematicians in the realm of integral equations I feel, however, that if Mr. Davis has actually succeeded in "throwing the bull" it is only in the sense as implied in the colloquial slang of his native America. In fact, we shall presently in some of our examples take up the challenge of Mr. Davis and show that the series he so curtly rejects can—by means of a simple transformation—be used on decidedly skew frequency distributions with even greater success than the Pearsonian curve types.

With these preliminary remarks we shall now proceed to give several examples of the application of the Gram or Laplacean—Charlier frequency series, employing either the method of moments or the method of least squares in the numerical determination of the constants, although preference will be given to the latter method in cases of appreciable skewness or excess.

It is, however, not our intention to go into details of the method of least squares and its relation to error laws, except in its connection with the problem of maximum and minimum. Any number of standard treatises are now available on the subject, however, to which we may refer interested readers.*

117. Charlier's Scheme of Computations.—The general formulae for the semi-invariants were given on page (192). In practical work it is, however, of importance to proceed along systematic lines and to furnish an automatic check for the correctness of the computations. Several systems facilitating such work have been proposed by various writers but the most simple and elegant is probably the one proposed by M. Charlier and which is shown in detail with the necessary control checks on the following pages. Charlier employs moments, while we in the following demonstration shall prefer the use of the semi-invariants.

* A particularly attractive presentation in English is found in David Brunt's *Combination of Observations* (Cambridge, 1918).

If we define the power sums of the relative frequencies $\varphi(x)$ by the relation $\mathrm{m}_r = \int_{-\infty}^{+\infty} x^r F(x)\,dx : \int_{-\infty}^{+\infty} F(x)dx$ for $r = 0,1,2,3 \ldots .$

we find that the expressions for the semi-invariants as given on page (192) may be written as follows:

$$\lambda_1 = m_1$$
$$\lambda_2 = m_2 - \ m_1^2$$
$$\lambda_3 = m_3 - 3m_2m_1 + 2m_1^3$$
$$\lambda_4 = m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4$$

. . . .

. . . .

The advantages of the Charlier scheme for the compuation of the semi-invariants lies in the fact that it furnishes an automatic check of the final results. If we expand the expression $(x+1)^4F(x)$ we have:

$$x^4F(x) + 4x^3F(x) + 6x^2F(x) + 4xF(x) + F(x) \text{ or}$$
$$\sum (x+1)^4 F(x) = s_4 + 4s_3 + 6s_2 + 4s_1 + s_0,$$

which serves as an independent control check of the computations. Moreover, another check is furnished by the relation

$$m_4 = \lambda + 4m_1\lambda_3 + 6m_1^2\lambda_2 + 3\lambda_2^2 + m_1^4.$$

In order to illustrate the scheme we chose the following age distribution of 1130 pensioned functionaries in a large American Public Utility corporation.

Ages	No. of Pensioners	Ages	No. of Pensioners
35–39	1	65–69	283
40-44	6	70–74	248
45–49	17	75–79	128
50–54	48	80–84	38
55–59	118	85–89	13
60–64	224	over 90	3

Choosing the age of 67 as a provisional origin the Charlier scheme is shown in detail on next page.

The computation below gives the numerical values of the frequency function which now may may be written as follows:

$$F(x) = 1130[\varphi_0(x) + .0258\varphi_3(x) + .0158\varphi_4(x)]$$

where

$$\varphi_0(x) = \frac{1}{1.624\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x+.0195}{1.6240}\right)^2}$$

Ages	x	$F(x)$	$xF(x)$	$x^2F(x)$	$x^3F(x)$	$x^4F(x)$	$(x+1)^4F(x)$
35–39	−6	1	6	36	216	1,296	625
40–44	−5	6	30	150	750	3,750	1,536
45–49	−4	17	68	272	1,088	4,352	1,377
50–54	−3	48	144	432	1,296	3,888	768
55–59	−2	118	236	472	944	1,888	118
60–66	−1	224	224	224	224	244	0
65–69	0	286	0	0	0	0	286
	Σ	700	708	1,586	4,518	15,418	4,710
70–74	+1	248	248	248	248	248	3,968
75–79	+2	128	256	512	1,024	2,048	10,368
80–84	+3	38	114	342	1,026	3,078	9,728
85–89	+4	13	52	208	832	3,328	8,125
90–94	+5	2	10	50	250	1,250	2,592
over 95	+6	1	6	36	216	1,296	2,401
	Σ	430	686	1,396	3,596	11,248	37,182
	s_r	1,130	−22	2,982	−922	26,666	41,892
	$m_r =$	1.0000	−.0195	2.6378	−.8156	23.5699	

$$\lambda_1 = m_1 = -.0195 \qquad m_2 = 2.6378 \qquad s_4 = 26{,}646$$
$$\lambda_1^2 = m_1^2 = .0004 \qquad -m_1^2 = -.0004 \qquad 4s_3 = -3{,}688$$
$$\lambda_1^3 = m_1^3 = .0000 \qquad \lambda_2 = 2.6374 = \sigma^2 \qquad 6s_2 = 17{,}892$$
$$\lambda_1^4 = m_1^4 = .0000 \qquad \sqrt{\lambda_2} = 1.6240 = \sigma \qquad 4s_1 = -\ 88$$
$$4.2831 = \sigma^3 \qquad s_0 = 1{,}130$$
$$6.9558 = \sigma^4 \qquad 41{,}892$$

$$m_2m_1 = -.0513, \quad m_3m_1 = .0159 \quad m_2^2 = 6.9580, \quad m_2m_1^2 = .0010$$

$$m_3 = -.8156 \qquad m_4 = 23.5699 \qquad \lambda_4 = 2.6450$$
$$-3m_2m_1 = .1539 \qquad -4m_3m_1 = -\ .0636 \qquad 4m_1\lambda_3 = .0516$$
$$2m_1^3 = .0000 \qquad -3m_2^2 = -20.8740 \qquad 6m_1^2\lambda_2 = 0060$$
$$\lambda_3 = -.0017 \qquad 12m_2m_1^2 = .0127 \qquad 3\lambda_2^2 = 20.8677$$
$$-6m_1^4 = .0000 \qquad m_1^4 = .0000$$
$$\lambda_4 = 2.6450 \qquad 23.5703 = m_4$$

$$c_3 = \lambda_3 : \sigma^3 = -.1545 \qquad c_4 = \lambda_4 : \sigma^4 = .3803$$
$$-c_3 : 3! = .0258 \qquad c_4 : 4! = +.0158$$

118. Comparison Between Observed Data and Theoretical Values.—The next step is now to work out the numerical values of $F(x)$ for various values of x and compare such values with the ones originally observed. This process is shown in detail in the following scheme:

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Obs.
x	$x-\lambda_1$	$(x-\lambda_1):\sigma$	$\varphi_0(z)$	$\varphi_3(z)$	$\varphi_4(z)$					
−7	−6.98	−4.300	.0001	+.0058	+.0170	+.0001	+.0003	.0005		
−6	5.98	3.682	.0005	.0176	.0479	.0015	.0008	.0028	2	1
−5	4 98	3.067	.0036	.0710	.1267	.0018	.0020	.0074	5	6
−4	3 98	2.451	.0198	.1458	+.0602	.0038	+.0009	.0245	17	17
−3	2.98	1.835	.0741	+.0500	−.4345	+.0013	−.0068	0686	48	48
−2	1.98	1.219	.1897	−.3502	−.7036	−.0090	.0111	.1696	118	118
−1	−0.98	−0.603	.3326	−.5287	+.3160	−.0136	−.0050	.3140	219	224
0	+0.02	+0.012	.3989	+.0143	1.1963	+.0004	+.0189	.4182	291	286
+1	1.02	0.628	.3273	.5359	+.2584	.0138	+.0041	.3452	241	248
+2	2.02	1.244	.1835	+.3325	−.7157	+.0086	−.0113	.1808	126	128
+3	3.02	1.860	.0707	−.0605	−.4094	−.0015	−.0065	.0627	44	38
+4	4.02	2.475	.0186	.1443	+.0703	.0037	+.0011	.0212	15	13
+5	5.02	3.091	.0034	.0680	.1241	.0018	.0020	.0036	3	2
+6	6.02	3.707	.0004	.0165	.0456	.0004	.0007	.0007	1	1
+7	+7.02	+4.322	.0001	−.0050	.0162	−.0001	+.0003	.0003		

Column (1) gives the values of the variate x reckoned from the provisional origin, or the centre of the age interval 65–69. (2) is x less the first semi-invariant, whereby the origin is shifted to the mean or λ. Column (3) represents the final linear transformation: $z=(x-\lambda_1):\sigma$.

Columns (4), (5) and (6) are copied directly from the standard tables of JØrgensen or Charlier. Column (7) is (5) multiplied by 0.0258 or the product $-[c_3\varphi_3(z)]:3!$, while (8) is $[c_4\varphi_4(z)]:4!$.

Column (9) is the sum of (4), (7) and (8) If we now distribute the area $N=s_0$ or 1130 PRO RATA according to (9), we finally reach the theoretical frequency distribution expressed in 5-year age intervals and shown in column (10) alongside which we have inserted the originally observed values. Evidently the fit is satisfactory. It will be noted that the final frequency series is expressed in units of 5-year age intervals. This, however, is only a formal representation. By subdividing the unit intervals of column (1) in 5 equal parts, and by computing all the other columns accordingly, we get the theoretical frequency series expressed in single year age intervals.

119. The Principle of Method of Least Squares.—The following paragraph purports to give a brief exposition of the determination of the coefficients in the Gram or Laplacean—Charlier series in the sense of the method of least squares as a strict problem of maxima and minima, wholly independent of the connection between the method of least squares and the error laws of precision measurements.*

*In the following demonstration I am adhering to the brief and lucid exposition of the Argentinean actuary, U. Broggi, in his exellent *Traite d'Assurances sur la Vie.*

The simple problem in maxima and minima which forms the fundamental basis of the method of least squares is the following: Let m unknown quantities be determined by observations in such a manner that they are not observed directly but enter into certain *known* functional relations, $f_i(x_1, x_2, x_3, \quad . \; . \; . \; . \; x_m)$, containing the unknown independent variables, $x_1, x_2, x_3, \ldots x_m$. Let furthermore the number of observations on such functional relations be n (where n is greater than m). The problem is then to determine the most plausible system of the values of the unknown x's from the observed system.

$$
\begin{aligned}
f_1\,(x_1, x_2, x_3, \; . \; . \; . \; x_m) &= o_1 \\
f_2\,(x_1, x_2, x_3, \; . \; . \; . \; x_m) &= o_2 \\
. \quad . \quad . \quad . \quad . \quad . \quad & \quad . \\
. \quad . \quad . \quad . \quad . \quad . \quad & \quad . \\
f_n\,(x_1, x_2, x_3, \; . \; . \; . \; x_m) &= o_n
\end{aligned}
$$

when $f_1, f_2, \; . \; . \; . \; f_n$ are the known functional relations and $o_1, o_2, \; . \; . \; . \; o_n$ their observed values. Such equations are known as *observation equations.*

In order to further simplify our problem we shall also assume that

1 All the equations of the system have the same weight, and

2 All the equations are reduced to linear form.

By these assumptions the problem is reduced to find m inknowns from n linear equations.

$$
\begin{aligned}
a_1\,x_1 + b_1\,x_2 + \; . \; . \; . \quad &= o_1 \\
a_2\,x_1 + b_2\,x_2 + \; . \; . \; . \quad &= o_2 \\
a_3\,x_1 + b_3\,x_2 + \; . \; . \; . \quad &= o_3 \\
. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad & \quad . \\
. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad & \quad . \\
a_n\,x_1 + b_n\,x_2 + \; . \; . \; . \quad &= o_n
\end{aligned}
$$

Since n is greater than m we find the problem over-determined, and we therefore seek to determine the unknown quantites, $x_1, x_2, \; . \; . \; . x_m$ in such a way that the sum of the squares of the differences between the functional relations and the observed values, o becomes a minimum. This implies that the expression

$$\sum_{i=1}^{i=n}(a_i x_1 + b_i x_2 + \; . \; . \; . \; - o_i)^2 = \psi(x_1, x_2, \; . \; . \; . \; x_m)$$

must be a minimum or the simultaneous existence of the equa tions.

$$\frac{d\psi}{dx_1}=0,\ \frac{d\psi}{dx_2}=0,\ \ldots\ \frac{d\psi}{dx_m}=0 \qquad (I)$$

If we now introduce the following notation

$$a_i x_1+b_i x_2+\ \ldots\ -o_i=\lambda_i \text{ for } i=1_1\,2_1\,3,\ \ldots\ n,$$

The m equations in the above system (I) evidently take on the following form

$$\begin{array}{l} \lambda_1 a_1+\lambda_2 a_2+\ \ldots\ +\lambda_n a_n=0 \\ \lambda_1 b_1+\lambda_2 b_2+\ \ldots\ +\lambda_n b_n=0 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{array}$$

If we now again re-substitute the expressions for λ in terms of the linear relations

$$a_i x_1+b_i x_2+\ \ldots\ o_i=\lambda_i, \text{ for } i=1, 2, 3, \ \ldots\ n,$$

and collect the coefficients of $x_1, x_2, \ \ldots\ x_n$, these equations may be expressed in the following symbolical form:

$$\begin{array}{l} [aa]x_1 + [ab]x_2 + \ldots\ldots - [ao] = 0 \\ [ab]x_1 + [bb]x_2 + \ldots\ldots - [bo] = 0 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ [ak]x_1 + [bk]x_2 + \ldots\ldots + [kk]x_m - [ko] = 0 \end{array}$$

where $[aa] = a_1^2 + a_2^2 + \ldots\ldots$
$[ab] = a_1 b_1 + a_2 b_2 + \ldots\ldots$

is the Gaussian notation for the homogenous sum products.

The above equations are known as *normal equations*, and it is readily seen that there is one normal equation corresponding to each unknown. Our problem is therefore reduced to the solution of a system of simultaneous linear equations of m unknowns. If m is a small number, or, what amounts to the same thing, there are only two or three unknowns the solution can be carried on by simple algebraic methods or determinants. If the number of unknowns is large these methods become very laborious and impractical. It is one of the achievements of the great German mathematician, Gauss, to have given us a method of solution which reduces this labor to a minimum and

which proceeds along well defined systematic and practical lines. The method is known as the Gaussian algorithmus of successive elimination.

120. Gauss' Solution of Normal Equations.—For the sake of simplicity we shall limit ourselves to a system of four normal equations of the form

$$
\begin{aligned}
[aa]x_1 + [ab]x_2 + [ac]x_3 + [ad]x_4 - [ao] &= 0 \\
[ab]x_1 + [bb]x_2 + [bc]x_3 + [bd]x_4 - [bo] &= 0 \\
[ac]x_1 + [bc]x_2 + [cc]x_3 + [cd]x_4 - [co] &= 0 \\
[ad]x_1 + [bd]x_2 + [cd]x_3 + [dd]x_4 - [do] &= 0
\end{aligned}
$$

The generalization to an arbitrary number of unknowns offers no difficulties, however.

On account of their symmetrical form the above equations may also be written in the more convenient form, viz.:

$$
\begin{aligned}
[aa]x_1 + [ab]x_2 + [acx_3 + [ad]x_4 - [ao] &= 0 \\
[bb]x_2 + [bc]x_3 + [bd]x_4 - [bo] &= 0 \\
[cc]x_3 + [cd]x_4 - [co] &= 0 \\
[dd]x_4 - [do] &= 0
\end{aligned}
$$

From the first equation we find

$$x_1 = \frac{[ao]}{[aa]} - \frac{[ab]}{[aa]}x_2 - \frac{[ac]}{[aa]}x_3 - \frac{[ad]}{[aa]}x_4$$

Substituting this value in the following equations and by the introduction of the new symbol

$$[ik] - \frac{[ai]}{[aa]}[ak] = [ik.1]$$

we now obtain a new system of equations of a lower order and of the form

$$
\begin{aligned}
[bb.1]x_2 + [bc.1]x_3 + [bd.1]x_4 - [bo.1] &= 0 \\
[cc.1]x_3 + [cd.1]x_4 - [co.1] &= 0 \\
[dd.1]x_4 - [do.1] &= 0
\end{aligned}
$$

Solving for x_2 we have

$$x_2 = \frac{[bo.1]}{[bb.1]} - \frac{[bc.1]}{[bb.1]}x_3 - \frac{[bd.1]}{[bb.1]}x_4$$

Substituting in the following equations and writing

$$[ik.1] - \frac{[bc.1]}{[bb.1]}[bk.1] = [ik.2]$$

we have

$$\begin{aligned} [cc.2]x_3 + [cd.2]x_4 &= [co.2] \\ [dd.2]x_4 &= [do.2], \text{ or} \end{aligned}$$

$$x_3 = \frac{[co.2]}{[cc.2]} - \frac{[cd.2]}{[cc.2]}x_4$$

Moreover, by writing

$$[ik.2] - [ci.2]\frac{[ck.2]}{[cc.2]} = [ik.3]., \text{ we have finally}$$

$$[dd.3]x_4 = [do.3]$$

This gives us the final reduced normal equation of the lowest order. By successive substitution we therefore have:

$$x_4 = \frac{[do.3]}{[dd.3]}$$

$$x_3 = \frac{[co.2]}{[cc.2]} - \frac{[cd.2]}{[cc.2]}x_4$$

$$x_2 = \frac{[bo.1]}{[bb.1]} - \frac{[bc.1]}{[bb.1]}x_3 - \frac{[bd.1]}{[bb.1]}x_4$$

$$x_1 = \frac{[ao]}{[aa]} - \frac{[ab]}{[aa]}x_2 - \frac{[ac]}{[aa]}x_3 - \frac{[ad]}{[aa]}x_4$$

as the ultimate solution of the unknowns.

121. Arithmetical Application of Method.—The example in paragraph 117 gave an illustration of the application of the method of moments. As previously stated this method works quite well in cases of moderate skewness, but is less successful in extremely skew curves and where the excess is large. We shall now give an illustration of the calculation of the parameters by the method of least squares. The example we choose is the well-known statistical series by the distinguished Dutch

botanist, deVries, on the number of petal flowers in *Ranunculus Bulbosus*.* This is also one of the classical examples of Karl Pearson in his celebrated original memoirs on skew variation. Although the observations of deVries lend themselves more readily to the method of logarithmic transformation, which we shall discuss in a following chapter, we have deliberately chosen to use it here for two specific reasons Firstly it is a most striking illustration in refutation of the incautious criticism of the Gram-Charlier series by the aforementioned Mr. Davis. Secondly (and this is the more important reason) it offers an excellent drill for the student in the practical applications of the method of least squares because it gives in a very brief compass all the essential arithmetical details. The observations of deVries are as follows:

No. of Petals	x	$F(x) = o_x$
5	0	133
6	1	55
7	2	23
8	3	7
9	4	2
10	5	2

where $F(x)$ denotes the absolute frequencies. The observed frequency distribution is well nigh as skew as it can be and represents in fact a one-sided curve, and should therefore—if the statement by Mr. Davis is correct—show an absolute defiance to a graduation by the Gram-Charlier series.

The process we shall use in the attempted mathematical representation of the above series is a combination of the method of semi-invariants and the method of least squares. Following Thiele's advice we determine the first two semi-invariants in the generating function directly from the observations while the coefficients of this function and its derivatives are determined by the least square method.

Choosing the provisional origin at 5, we obtain the following values for the crude moments.

*Undoubtedly many readers will think that I have spent an unusual heavy amount of arithmetic on such a simple example. This criticism is true, and in actual curve fitting practice we would of course resort to a logarithmic transformation. The example is in this particular instance chosen as a drill for the student. I may, however remark that if one were to use Personian curves the arithmetical work would be even more formidable than through the application of least squares; because we would have to resort to mechanical quadrature formulas in order to compute the areas in Pearson's curves.

$s_0=222$, $s_1=140$, $s_2=292$, $s_3=806$, $s_4=2{,}752$, $s_5=10{,}790$, $s_6=46{,}072$, $s_7=207{,}226$, from which we find that

$\lambda_0=1$, $\lambda_1=0.631$, $\lambda_2=0.917$, $\lambda_3=1.644$, $\lambda_4=3.377$, $\lambda_5=5.972$, $\lambda_6=-2.911$, $\lambda_7=-122.638$.

All these semi-invariants with the exception of the two first are however, so greatly influenced by random sampling in the small observation series that it is hopeless to use them in the determination of the constants in the Gram-Charlier series. In fact an actual calculation does not give a very good result beyond that of a first rough approximation. The generating function, on the other hand, may be expressed by the aid of the two first semi-invariants as follows:

$$\varphi_0(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2:2},$$

where z is given by the linear transformation:

$$z=(x-0.631):0.9576. \qquad (\sqrt{\lambda_2}=0.9576).$$

We now propose to express the observed function $F(x)$ or $\varphi(z)$ by a Gram-Charlier series of the form:

$$F(x)=\varphi(z)=k_0\varphi_0(z)+k_3\varphi_3(z)+k_4\varphi_4(z)+\ .\ .\ .\ .\ .\ k_7\varphi_7(z).$$

In this equation we know the values of the generating function and its derivatives for various values of the variate z as found in the tables of Jørgensen and Charlier, while the quantities k are unknowns. On the other hand we know 6 specific values of $F(x)$ as directly observed in deVries's observation series. We are thus dealing with a system of typical linear observation equations of the forms described in paragraphs 119 and 120 and which lend themselves so admirably to the treatment by the method of least squares.

From the above linear relation between x and z we can directly compute the following table for the transformed variate z.

x	z
0	-0.688
1	$+0.402$
2	$+1.493$
3	$+2.583$
4	$+3.674$
5	$+4.764$

The numerical values of $\varphi_0(z)$ and its derivatives as corresponding to the above values of z can be taken directly from the standard tables of Jørgensen and Charlier. We may therefore write down the following observation equations (a)

φ_0	φ_3	φ_4	φ_5	φ_6	o
$.3148k_0$	$-.5472k_3$	$+.1207k_4$	$+2.2728k_5$	$+.9591k_6$	$-133=0$
$.3679k_0$	$+.4198k_3$	$+.7566k_4$	$-1.9836k_5$	$-2.9860k_6$	$-55=0$
$.1308k_0$	$+.1506k_3$	$-.7073k_4$	$+.4540k_5$	$+2.8600k_6$	$-23=0$
$.0145k_0$	$-.1346k_3$	$+.1062k_4$	$+.2642k_5$	$-1.2130k_6$	$-7=0$
$.0005k_0$	$-.0180k_3$	$+.0486k_4$	$-.1070k_5$	$+.1482k_6$	$-2=0$
$.0001k_0$	$-.0005k_3$	$+.0020k_4$	$-.0043k_5$	$+.0540k_6$	$-2=0$

for which we now propose to determine the unknown values of k by the least square method.

While this method may of course be applied directly to the above data, it will generally be found of advantage to start with some approximate values of the k's. It is found in practice that this approximate step saves considerable labour in the formation and ultimate solution of the normal equations.

Although the first approximation in the case of numerous unknowns must be in the nature of a more or less shrewd guess, which facility can only be attained by constant practice in routine mathematical computing, we are, however, in this specific instance able to tell something about the nature of the coefficients from purely *a priori* considerations. We know for instance from the form of the Gram-Charlier series that the coefficient k_0 of the generating function must be nearly equal to the area of the curve, which in this particular instance is 222. Moreover, a mere glance at the observed series tells us that it has a decidedly large skewness in negative direction from the mean coupled with a tendency of being "top heavy," indicating positive excess. We can therefore assume as a first approximation that the coefficients of the derivatives of uneven order are negative and the coefficients of derivatives of even order are positive. Again, it is also seen that the coefficients of the derivatives of higher order than the fourth must be relatively small in comparison with the coefficients of the derivatives of lower order, otherwise the series would not be rapidly convergent.

From such purely common sense *a priori* considerations we therefore guess the following first approximations, viz.:

$$k'_0=222,\ k'_3=-100,\ k'_4=20,\ k'_5=-5,\ k'_6=5.$$

The probable values of the various k's may be written as

$$k_i = r_i k^1_i \text{ for } i = 0, 3, 4, 5 \text{ and } 6,$$

and our problem is therefore to find the correction factor r_i with which the approximate value k^1_i must be multiplied so as to give k_i.

Applying the various values of k^1_i to the original observation equations in (a) we obtain the following schedule for the numerical factors of r_i.

a	b	c	d	e	o	s
699	+547	+ 24	−144	+ 48	−1,330	−126
817	−420	+151	+ 99	−149	− 550	− 52
290	−151	−141	− 23	+143	− 230	−112
32	+135	+ 21	− 13	− 61	− 70	+ 44
1	+ 18	+ 10	+ 5	+ 7	−. 30	+ 11
0	+ 0	+ 0	+ 0	+ 3	− 10	− 7
1839	+129	+ 65	− 46	− 9	−2,220	−242

where the additional control column s serves as a check.

The subsequent formation of the various sum-products and normal equations is shown in the following schedules together with the s columns as a check.

aa	ab	ac	ad	ae	ao	as
488,601	+382,353	+ 16,776	− 79,686	+ 33,552	− 929,670	− 88,074
667,489	−343,140	+123,367	+ 80,883	−121,733	− 449,350	− 42,484
84,100	− 43,790	− 40,890	− 6,670	+ 41,470	− 66,700	− 32,480
1,024	+ 4,320	+ 672	− 416	− 1,952	− 2,240	+ 1,408
1	+ 18	+ 10	+ 5	+ 7	− 30	+ 11
0	+ 0	+ 0	+ 0	+ 0	− 0	0
1,241,215	− 239	+ 99,935	− 5,884	− 48,656	−1,447,990	−161,619

bb	bc	bd	be	bo	bs
299,209	+ 13,128	− 63,258	+ 26,256	− 727,510	− 68,922
176,400	− 63,420	− 41,580	+ 62,580	+ 231,000	+ 21,840
22,801	+ 21,291	+ 3,473	− 21,593	+ 34,730	+ 16,912
18,225	+ 2,835	− 1,755	− 8,235	− 9,450	+ 5,940
324	+ 180	+ 90	+ 126	− 540	+ 198
0	+ 0	+ 0	+ 0	− 0	+ 0
516,959	− 25,986	−102,130	+ 59,134	− 471,770	− 24,032

cc	cd	ce	co	cs
576	− 2,736	+ 1,152	− 31,920	− 3,024
22,801	+ 14,949	− 22,499	− 83,050	− 7,852
19,881	+ 3,243	− 20,163	+ 32,430	+ 15,792
441	− 273	− 1,281	− 1,470	+ 924
100	+ 50	+ 70	− 300	+ 110
0	+ 0	+ 0	− 0	− 0
43,799	+ 15,233	− 42,721	− 84,310	+ 5,950

dd		*de*		*do*		*ds*
12,996	−	5,472	+	151,620	+	14,364
9,801	−	14,751	−	54,450	−	5,148
529	−	3,289	+	5,290	+	2,576
169	+	793	+	910	−	572
25	+	35	−	150	+	55
0	+	0	+	0	+	0
23,520	−	22,684	+	103,220	+	11,275

ee		*eo*		*es*
2,304	−	63,840	−	6,048
22,201	+	81,950	+	7,748
20,449	−	32,890	−	16,016
3,721	+	4,270	−	2,684
49	−	210	+	77
9	−	30	−	21
48,733	−	10,750	−	16,944

We may now write the normal equations in schedule form as follows:

ORIGINAL NORMAL EQUATIONS

(*a*)	1,241,215	− 239	+ 99935	− 5884	− 48656	−1447990
(1)		+ *0*	− *19*	+ *1*	+ *9*	+ *.278*
(*b*)		+516959	− 25986	−102130	+ 59134	− 471770
(2)			+ *8046*	− *474*	− *3917*	− *116582*
(*c*)			+ 43799	+ 15233	− 42721	− 84310
(3)				+ *28*	+ *231*	+ *6865*
(*d*)				+ 23520	− 22684	+ 103220
(4)					+ *1907*	+ *56761*
(*e*)					+ 48733	− 10750
(5)		*−.00019*	*+.08051*	*−.00474*	*−.03920*	*−1.16659*

The sum-products from the observation equations are shown in the rows marked (*a*), (*b*), (*c*), (*d*) and (*e*). The row marked (5) and printed in italics is formed by dividing each of th figures in row (*a*) with 1,241,215. The row marked (1) contains the products of the figures in row (*a*) multiplied with the factor .00019. Row (2) is the products of the factor 0.08051 and the figures in row (*a*), while row (3) is the product of the factor −0.00474 and the figures in row (*a*). The products in row (4) are formed in the same manner by means of the factor −0.03920.

We next subtract row (1) from row (*b*), row (2) from row (*c*), row (3) from row (*d*), and so forth, which results in the following schedule, which is known as the first *reduction equation*.

FIRST REDUCTION EQUATION

(a)	+516959	− 25967	−102131	+ 59125	− 472048
(1)		+ *1304*	+ *5130*	− *2970*	+ *23711*
(b)		+ 35753	+ 15707	− 38804	+ 32272
(2)			+ *20177*	− *11681*	+ *93258*
(c)			+ 23492	− 22915	+ 96355
(3)				+ *6762*	− *53988*
(d)				+ 46826	− 67511
(4)		*−.05023*	*−.19756*	*+.11437*	*− .91313*

The above equations are treated in a similar manner as the original normal equations, and we have therefore the 2d reduction equation of the form:

SECOND REDUCTION EQUATION

(a)	+ 34451	+ 10577	− 35834	+ 8561
(1)		+ *3247*	− *11002*	+ *2628*
(b)		+ 3315	− 11234	+ 3097
(2)			+ *37273*	− *8905*
(c)			+ 40064	− 13523
(3)		*+.30702*	*−1.04014*	*+ .24850*

THIRD REDUCTION EQUATION

(a)	+ 68	− 232	+ 469
(1)		+ *791*	− *1600*
(b)		+ 2791	− 4618
(2)		*−3.41170*	*+6.89706*

FOURTH REDUCTION EQUATION

+ 2000	− 3018

The solution for the unknown r's may now be shown as follows:

$$r_6 = 3018 : 2000 = 1.5090$$

$$r_5 = -6.8971 - (1.5090)(-3.4117) = -1.7488$$

$$r_4 = -0.2485 - (-1.7488)(0.3070) - (1.5090)(-1.0401) = 1.8580$$

$$r_3 = 0.9131 - (1.8580)(-0.0502) - (-1.7488)(-0.1976) - (1.5090)(0.1144) = 0.4884$$

$$r_0 = 1.1666 - (0.4884)(-0.0002) - (1.8580)(0.0805) - (-1.7488)(-0.0047) - (1.5090)(-0.0394) = 1.0679$$

From the above values of r and by means of the relation

$$k_i = r_i \, k_i^2 \text{ for } i = 0, 3, 4, 5 \text{ and } 6$$

we can easily determine the most probable values of k_i with which the original observation equations as shown on page (228) must be multiplied so as to satisfy the observed values of $F(x)$ in the sense as implied in the method of least squares.

This results in the following arrangement:

z	$k_0\varphi_0(z)$	$k_3\varphi_3(z)$	$k_4\varphi_4(z)$	$k_5\varphi_5(z)$	$k_6\varphi_6(z)$	$\Sigma k_i\varphi_i$	Obs.
−0.688	74.6	+26.7	+ 4.4	+19.7	+ 7.2	132.6	133
+0.402	87.2	−20.5	+28.1	−17.2	−22.4	55.2	55
1.493	31.0	− 7.3	−26.2	+ 3.9	+21.5	22.9	23
2.583	3.4	+ 6.6	+ 3.9	+ 2.3	− 9.1	7.1	7
3.674	0.2	+ 0.9	+ 1.8	− 0.9	+ 1.0	3.0	2
+4.764	0.0	+ 0.0	+ 0.1	+ 0.0	+ 0.0	0.1	2

The agreement between the calculated values and the originally observed series leaves evidently little to be desired in the way of a satisfactory "fit."

If we limited ourselves to three terms of the series and put

$$F(x) = \varphi(z) = \sum k_i \varphi_i(z) \text{ for } i = 0, 3 \text{ and } 4$$

and then determined k_0, k_3 and k_4 by the method of least squares the final result would be of the form:

$$\varphi(z) = 264.2\varphi_0(z) - 89.9\varphi_3(z) - 5.2\varphi_4(z),$$

for which the calculated values of the frequency function would be as follows:

x	$F(x)$	Obs.	Pearson
5	131.6	133	136.9
6	55.2	55	48.5
7	24.5	23	22.6
8	15.5	7	9.6
9	1.6	2	3.4
10	0.2	2	0.8

The fit is evidently not so close as when we use 6 terms, but it is by no means a poor fit and does not require nearly so much arithmetical work as the larger number of terms in the frequency series. In this connection it is of interest to compare the present graduation with the result reached by Pearson, which is also shown in the above table. Taken all in all there

is no doubt in my mind that the serial expansion gives far better results than the Pearsonian methods and does not entail nearly so much labour as these.

Note on Adjusted Moments.—The theoretical moments are given in the form of definite integrals while the observations always give us the moments on the form

$$s_r = \sum_{x=-\infty}^{x=+\infty} x^r \int_{x-\frac{1}{2}a}^{x+\frac{1}{2}a} \varphi(x)\,dx$$

where a is the class interval of the observations. In order to determine the semi-invariants it is, however, required to know the continuous moments

$$M_r = \int_{-\infty}^{+\infty} x^r \varphi(x)\,dx$$

The values of s being given in the form of finite sums and not as definite integrals are therefore subject to certain adjustments if we wish to express them as continuous moments. The necessary adjustments can, however, easily be performed by well-known formulas from the theory of mechanical quadrature if the frequency function and its derivatives vanish for $x = -\infty$ and $x = +\infty$. The English mathematician, Sheppard, has among others developed the following simple formulas for the transition from s to M:

$$M_0 = s_0,\ M_1 = s_1,\ M_2 = s_2 - \frac{a}{12}s_2,\ M_3 = s_3 - \frac{a^2}{4}s_1$$

$$M_4 = s_4 - \frac{a^2}{2}s_2 + \frac{7a^4}{240}s_0,\ M_5 = s_5 - \frac{5a^2}{6}s_3 + \frac{7a^2}{48}s_1$$

The Sheppard adjustments again emphasizes the fact that the method of moments works with curve areas instead of curve ordinates, which necessarily must lead to some sort of mechanical quadrature formula unless we are able to evaluate the indefinite integrals of the expressions for the frequency functions. If we use curve ordinates to calculate the specific numerical values of Pearson's frequency functions we are liable to encounter large errors. This fact is among other things pointed out by Caradog Jones who in mentioning the use of ordinates points out that "it must be remembered the resulting values are only a first approximation to the observed frequencies and a better series is obtained if, by using some good quadrature formula, we calculate the AREAS for the successive groups between the curve, the bounding ordinates, and the axis of x." This is one of the great drawbacks to the otherwise elegant Pearsonian types of frequency curves because it entails a large amount of arithmetical work to compute specific numerical values from the final formulas as determined by Pearson's curve types.

Any reader who will take the trouble to consult the original memoirs by Pearson and Elderton and the recently published treatise by Mr.

Caradog Jones will there find ample evidence of the large amount of tedious arithmetical work involved in the application of mechanical quadrature formulas. The recently suggested finite difference equation formulas by the American mathematican, Carver, while emphasizing the difficulty of applying the Pearson system, do not tend materially to shorten the arithmetical work very much, and Mr. Carver must in the final instance resort to mechanical quadrature.*

All these difficulties are, however, eliminated in the case of the determination of the frequency function in serial form by means of the method of least squares where we work equally well with ordinates as with areas. A further advantage of the Gram—Charlier expansion in series is found in the fact that standard tables of the generating function and its derivatives as well as the definite integrals of these functions have been published both by Bruhns, Charlier and Jorgensen. Speaking from a purely personal point of view I wish to state that through a long and varied experience in practical curve fitting to the most diverse kinds of statistical data I have had occasion to use both the Pearsonian and the Gram—Charlier type of curves, and while I fully recognize the theoretical elegance and apparent simplicity of the Pearson system, I feel nevertheless that from the point of view of the practical computer the older system as devised by the Scandinavian investigators is to be preferred in comparison with the methods advocated by the followers of the distinguished founder of the English Biometric school.

*Mr. Carver's able and interesting analysis by means of finite difference equations is, however, to a great extent anteceded by the much earlier Danish memoirs of Opperman and Gram where the finite difference equation methods are discussed.

CHAPTER XVI

LOGARITHMICALLY TRANSFORMED FREQUENCY FUNCTIONS

122. Transformation of the Variate.—While it is always possible to express all frequency curves by an expansion in Hermite polynomials, the numerical labor when carried on by the method of least squares often involves a large amount of arithmetical work if we wish to retain more than four or five terms of the series. Other methods lessening the arithmetical work and making the actual calculations comparatively simple have been offered by several authors and notably by Thiele, who in his works discusses several such methods. Among those we may mention the method of the so-called free functions and orthogonal substitution, the methods of correlates and the adjustment by elements. The chapters on these methods in Thiele's work are among some of the most important, but also some of the most difficult in the whole theory of observations and have not always been understood and appreciated by the mathematicians, chiefly on account of Thiele's peculiar style of writing. A close study of the Danish scholar's investigations is, however, well worth while, and Thiele's work along these lines may still in the future become as epochmaking in the theory of probability as some of the researches of the great Laplace. The theory of infinite determinants as used by M. Fredholm in the solution of integral equations is another powerful tool which offers great advantages in the way of rapid calculation. All these methods require, however, that the student must be thoroughly familiar with the difficult theory upon which such methods rest, and they have for this reason been omitted in an elementary work such as the present treatise.

We wish, however, to mention another method which in the majority of cases will make it possible to employ the Gram or Laplacean—Charlier curves in cases with extreme skewness or excess. We have here reference to the method of logarithmic transformation of the variate, x.

123. The General Theory of Transformation.—One of the simplest transformations is the previously mentioned linear transformation of the form $z=f(x)=ax+b$, by which we can make two constants, c_1 and c_2 vanish. Other transformations suggest themselves, however, such as $f(x)=ax^2+bx+c$, $f(x)=\sqrt{x}$, $f(x)=\log x$ and so forth. For this reason I propose to give a brief development of the general method of transformations of the statistical variates, mainly following the methods of Charlier and Jørgensen.

Stated in its most general form our problem is: If a frequency curve of a certain variate is given by $F(x)$ what will be the frequency curve of a certain function of x, say $f(x)$?

The equation of the frequency curve is $y=F(x)$, which means that $F(x)dx$ is the probability that x falls in the interval between $x-\frac{1}{2}dx$ and $x+\frac{1}{2}dx$. The probability that a new variate z after the transformation $z=f(x)$, or $\chi(z)=x$, falls in the interval $z-\frac{1}{2}dz$ and $z+\frac{1}{2}dz$ is therefore simply

$$F[\chi(z)]\chi'(z)dz=F(x)dx,$$

which gives in symbolic form the equation of the transformed frequency curve.

The frequency for $z=f(x)$ is of course the same as for x. The ordinates of the frequency curve, or rather the areas between corresponding ordinates, are therefore not changed, but the abcissa axis is replaced by $f(x)$. Equidistant intervals of x will therefore not as a rule—except in the linear transformation—correspond to equidistant intervals of $f(x)$.

If, for instance, the frequency curve $F(x)$ is the Laplacean normal curve

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2:2\sigma^2}$$

and if we let $z=f(x)=x^2$ or $x=\sqrt{z}$, we have evidently

$$F(z)=\frac{1}{\sigma\sqrt{2\pi}}\;\frac{e^{-z:2\sigma^2}}{2\sqrt{z}}$$

124. Logarithmic Transformation.—Of the various transformations the logarithmic is of special importance. It happens that even if the variate x forms an extremely skew frequency distribution its logarithms will be nearly normally distributed.

This fact was already noted by the eminent German psychologist, Fechner, and also mentioned by Bruhns in his *Kollektivmasslehre.* But neither Fechner nor Bruhns have given a satisfactory theoretical explanation of the transformation and have limited themselves to use it as a practical rule of thumb.

Thiele discusses the method under his adjustment by elements, but in a rather brief manner. The first satisfactory theory of logarithmic transformation seems to have been given first by Jørgensen and later on by Wicksell.* Jørgensen first begins with the transformation of the normal Laplacean frequency curve. Letting $z = \log x$ and bearing in mind that the frequency of x equals that of $\log x$ we have

$$z = f(x) = \log x, \text{ or } x = \chi(z) = e^z \text{ and } dx = e^z dz$$

The continuous power sums or moments of the rth order around an arbitrary origin take on the form

$$M_r = (n\sqrt{2\pi})^{-1}N\int_{-\infty}^{+\infty} x^r e^{-\frac{1}{2}\left(\frac{x-m}{n}\right)^2} dx = (n\sqrt{2\pi})^{-1}N\int_{0}^{\infty} x^r e^{\frac{1}{2}\left(\frac{\log x - m}{n}\right)^2} dx$$

$$= (n\sqrt{2\pi})^{-1}N\int_{-\infty}^{+\infty} e^{rz} e^{-\frac{1}{2}\left(\frac{z-m}{n}\right)^2} e^z dz$$

The change in the lower limit in the second integral from $-\infty$ to zero arises simply from the fact that the logarithm of zero equals minus infinity and the point $-\infty$ is thus by the transformation moved up to zero.

By a straightforward transformation (see appendix) we may write the above integral as

$$M_r = \frac{N}{\sqrt{2\pi}} e^{m(r+1)+\frac{1}{2}n^2(r+1)^2} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt = N e^{m(r+1)+\frac{1}{2}n^2(r+1)^2}$$

Changing from moments to semi-variants by means of the well-known relations

$$\lambda_0 = M_0$$

$$\lambda_1 = M_1 : M_0$$

$$\lambda_2 = (M_2 M_0 - M_1^2) : M_0^2$$

*The law of errors, leading to the geometric mean as the most probable value of the variate as discovered by Sir Donald McAlister in 1879 may, however, be considered as a forerunner of Jørgensen's work.

$$\lambda_3 = (M_3M_0^2 - 3M_2M_1M_0 + 2M_1^3) : M_0^3$$

$$\lambda_4 = (M_4M_0^3 - 4M_3M_1M_0^2 - 3M_2^2M_0^2 + 12M_2M_1^2M_0 - 6M_1^4) : M_0^4$$

. . . .

. . . .

we have

$$\lambda_0 = Ne^{m+\frac{1}{2}n^2}$$

$$\lambda_1 = e^{m+1.5n^2}$$

$$\lambda_2 = e^{2m+3n^2}(e^{n^2}-1)$$

$$\lambda_3 = e^{3m+4.5n^2}(e^{3n^2}-3e^{n^2}+2)$$

$$\lambda_4 = e^{4m+6n^2}(e^{6n^2}-4e^{3n^2}-3e^{2n^2}+12e^{n^2}-6)$$

. . . .

. . . .

These equations give the semi-invariants expressed in terms of m and n. On the other hand if we know the semi-invariants from statistical data or are able to determine these semi-invariants by *a priori* reasoning we may find the parameters m and n.

125. The Mathematical Zero.—A point which we must bear in mind is that the above semi-invariants on account of of the transformation are calculated around a zero point which corresponds to a fixed lower limit of the observations.

Very often the observations themselves indicate such a lower limit beyond which the frequencies of the variate vanish In the case of persons engaged in factory work there is in most countries a well-defined legal age limit below which it is illegal to employ persons for work. Another example is offered in the number of alpha particles radiated from certain radioactive metals. Since the number of particles radiated in a certain interval of time must either be zero or a whole positive number it is evident that—1 must be the lower limit because we can have no negative radiations. Analogous limits exist in the age limit for divorces and in the amount of moneys assessed in the way of income tax.

The lower limit allows, however, of a more exact mathematical determination by means of the following simple considerations. It is evident that this lower limit must fall below

the mean value of the frequency curve. Let us suppose that it is located at a certain point, a, at a distance of η units from the mean $M=\lambda_1(x)-\eta=a$; and let us furthermore as a beginning place the origin at $\lambda_1(x)$, in which case λ_1 of course equals zero. By shifting the origin to a, which implies a translation of η units in negative direction, the original variate (x) is transformed into $x+\eta$, and λ_1 will now equal η while the semi-invariants of higher order remain the same as before the transformation because of the well known relation

$$\lambda_r(x-\eta)=\lambda_r(x) \text{ for } r>1$$

We may therefore write the previously given relations between the λ's and m and n as follows:

$$\lambda_1-a=\eta=e^{m+1.5n^2}$$

$$\lambda_2=\eta^2(e^{n^2}-1) \text{ or } e^{n^2}=1+\lambda_2:\eta^2$$

$$\lambda_3=\eta^3\left[\left(\frac{\lambda_2}{\eta^2}+1\right)^3-3\left(\frac{\lambda_2}{\eta^2}+1\right)+2\right]=\eta^3\left[\frac{\lambda_2^3}{\eta^6}+\frac{3\lambda_2^2}{\eta^4}\right]$$

which reduces to $\lambda_3\eta^3-3\lambda_2^2\eta^2-\lambda_2^3=0$.

The solution of this cubic equation which has one real and two imaginary roots gives us the value of η or $\lambda_1 - a$ and thus determines the mathematical zero or lower limit. We have in fact:

$$n^2=\log(1+\lambda_2:\eta^2) \text{ and}$$

$$m=\log\eta-1.5n^2, \text{ while}$$

$$N=\lambda_0:e^{m+\frac{1}{2}n^2}$$

126. Logarithmically Transformed Frequency Series.—We have already shown that the generalized frequency curve could be written as

$$F(x)=c_0\varphi_0(x)-\frac{c_1\varphi_1(x)}{1!}+\frac{c_2\varphi_2(x)}{2!}-\frac{c_3\varphi_3(x)}{3!}+\cdots\cdots$$

where the Laplacean probability function

$$\varphi_0(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-M)^2}{2\sigma^2}}$$

is the generating function with M and σ as its parameters.

The suggestion now immediately arises to use an analogous series in the case of the logarithmic transformation. In this case the frequency curve, $F(x)$, with a lower limit would be expressed as follows:

$$F(x)=k_0\Phi_0(x)-\frac{k_1\Phi_1(x)}{1!}+\frac{k_2\Phi_2(x)}{2!}-\frac{k_3\Phi_3(x)}{3!}+\cdots\cdots$$

while the generating function now is

$$\Phi_0(x)=\frac{1}{n\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\log x-m}{n}\right]^2}$$

where m and n are the parameters.

Using the usual definition of semi-invariants we then have

$$s_0e^{\frac{\lambda_1\omega}{1!}+\frac{\lambda\omega^2}{2!}+\frac{\lambda_3\omega^3}{3!}+\cdots\cdots}=s_0+\frac{s_1\omega}{1!}+\frac{s_2\omega^2}{2!}+\frac{s_3\omega^3}{3!}+\cdots\cdots$$

$$=\int_0^\infty e^{x\omega}[k_0\Phi_0(x)-\frac{k_1\Phi_1(x)}{1!}+\frac{k_2\Phi_2(x)}{2!}-\frac{k_3\Phi_3(x)}{3!}+\cdots,.]dx.$$

The general term on the right hand side integral is of the form

$$(-1)^s k_s{:}s!\int_0^\infty e^{x\omega}\Phi_s(x)\,dx$$

where the integral may be evaluted by partial integration as follows:

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx=e^{x\omega}\Phi_{s-1}(x)\Big]_0^\infty-\omega\int_0^\infty e^{x\omega}\Phi_{s-1}(x)\,dx$$

Since both $\Phi_0(x)$ and all its derivatives are supposed to vanish for $x=0$ and $x=\infty$ the first term to the right becomes zero and

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx=-\omega\int_0^\infty e^{x\omega}\Phi_{s-1}(x)\,dx$$

By successive integrations we then obtain the following recursion formula

$$(-\omega)^1\int_0^\infty e^{x\omega}\Phi_{s-1}(x)\,dx = (-\omega)^2\int_0^\infty e^{x\omega}\Phi_{s-2}(x)\,dx$$

$$(-\omega)^2\int_0^\infty e^{x\omega}\Phi_{s-2}(x)\,dx = (-\omega)^3\int_0^\infty e^{x\omega}\Phi_{s-3}(x)\,dx$$

. . . .

. . . .

$$(-\omega^{s-1}\int_0^\infty e^{x\omega}\Phi_1(x)\,dx = (-\omega)^s\int_0^\infty e^{x\omega}\Phi_0(x)\,dx$$

Or finally

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx = (-\omega)^s\int_0^\infty e^{x\omega}\Phi_0(x)\,dx$$

Expanding $e^{x\omega}$ in a power series we have

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx = \frac{(-\omega)^s}{n\sqrt{2\pi}}\int_0^\infty [1+x\omega+\frac{x^2\omega^2}{2!}+\frac{x^3\omega^3}{3!}+\ldots]e^{-\frac{1}{2}\left[\frac{\log x-m}{n}\right]^2}dx$$

The general term in this expansion is of the form

$$\frac{(-\omega)^s}{n\sqrt{2\pi}}\frac{\omega^r}{r!}\int_0^\infty x^r e^{-\frac{1}{2}\left[\frac{\log x-m}{n}\right]^2}dx$$

which according to the formulas given on page (237) reduces to:

$$(-\omega)^s e^{m(r+1)+\frac{1}{2}n^2(r+1)^2}\omega^r : r\,!$$

Hence we may write

$$\int_0^\infty e^{x\omega}\Phi_s(x)\,dx = (-\omega)^s\sum_{r=0}^{r=\infty} e^{m(r+1)+\frac{1}{2}n^2(r+1)^2}\omega^r : r!$$

Consequently the relation between the semi-invariants and the frequency function

$$F(x) = k_0\Phi_0(x) - \frac{k_1}{1!}\Phi_1(x) + \frac{k_2}{2!}\Phi_2(x) - \frac{k_3}{3!}\Phi_3(x) + \ldots.$$

can be expressed by the following recursion formula

$$s_0 e^{\frac{\lambda_1\omega}{1!}+\frac{\lambda_2\omega^2}{2!}+\frac{\lambda_3\omega^3}{3!}+\cdots} = s_0+\frac{s_1\omega}{1!}+\frac{s_2\omega^2}{2!}+\frac{s_3\omega^3}{3!}+\cdots$$

$$=\sum_{v=0}^{v=\infty} s_v\frac{\omega^v}{v!}=\sum_{s=0}^{s=\infty}\frac{k_s}{s!}\omega^s\sum_{r=0}^{r=\infty}e^{m(r+1)+\frac{1}{2}n^2(r+1)^2}\omega^r:r!$$

The constants k are here expressed in terms of the unadjusted moments or power sums, s. It is readily seen that the Sheppard corrections for adjusted moments, M, also apply in this case. We are, therefore, able to write down the values of the k's from the above recursion formula in the following manner

$$M_0=k_0e^{m+\frac{1}{2}n^2}$$
$$M_1=k_1e^{m+\frac{1}{2}n^2}+k_0e^{2m+2n^2}$$
$$M_2=k_2e^{m+\frac{1}{2}n^2}+2k_1e^{2m+2n^2}+k_0e^{3m+4.5n^2}$$
$$M_3=k_3e^{m+\frac{1}{2}n^2}+3k_2e^{2m+2n^2}+3k_1e^{3m+4.5n^2}+k_0e^{4m+8n^2}$$
$$M_4=k_4e^{m+\frac{1}{2}n^2}+4k_3e^{2m+2n^2}+6k_2e^{3m+4.5n^2}+4k_1e^{4m+8n^2}+k_0e^{5m+12.5n^2}$$

It is easy to see that it is not possible to determine the generating function's parameters m and n from the observations. These parameters like M and σ in the case of the Laplacean normal probability curve must be chosen arbitrarily. If m and n are selected so as to make k_1 and k_2 vanish we have

$$M_0=k_0e^{m+\frac{1}{2}n^2}$$
$$M_1=k_0e^{2m+2n^2}$$
$$M_2=k_0e^{3m+4.5n^2}$$

the solution of which gives

$$e^{n^2}=\frac{M_0M_2}{M_1^2},\quad e^{2m}=\frac{M_1^8}{M_0^5M_2^3},\quad k_0=\frac{M_0^3M_2}{M_1^3}$$

while

$$k_3e^{m+\frac{1}{2}n^2}=M_3-M_0e^{3m+7.5n^2}$$
$$k_4e^{m+\frac{1}{2}n^2}=M_4-4M_3e^{m+1.5n^2}-M_0e^{4m+9n^2}(e^{3n^2}-4)$$

This theory requires the computation of a set of tables of the generating function

$$\Phi_0(x) = \frac{1}{n\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\log x - m}{n}\right]^2}$$

and its derivatives. For $\Phi_0(x)$ itself we may of course use the ordinary tables for the normal curve $\varphi_0(z)$ when we consider

$$z = \frac{\log x - m}{n}$$

I have calculated a set of tables of the derivatives of $\Phi_0(x)$ and hope to be able to publish the manuscript thereof in the second volume of this treatise.

127. Parameters Determined by Least Squares.—The above development is based upon the theory of functions and the theory of definite integrals. We shall now see how the same problem may be attacked by the method of least squares after we have determined by the usual method of moments the values of m and n in the generating function $\varphi_0(z)$.

Viewed from this point of vantage our problem may be stated as follows:

Given an arbitrary frequency distribution, of the variate z with $z = (\log x - m):n$ and where x is reckoned from a zero point or origin a, which is situated η units below the mean and defined by the relation

$$\eta^3\lambda_3 - 3\eta^2\lambda_2^2 = \lambda_2^3, \text{ where } \eta = \lambda_1 - a;$$

to develop $F(z)$ into a frequency series of the form

$$F(z) = k_0\varphi_0(z) + k_3\varphi_3(z) + k_4\varphi_4(z) + \ldots, . + k_n\varphi_n(z),$$

where the k's must be determined in such a way that the expression

$$\sum_{i=0}^{i=n} k_i\varphi_i(z)$$

gives the best approximation to $F(z)$ in the sense of the method of least squares.

Stated in this form the frequency function is reduced to the ordinary series of Gram or the A type of the Charlier series, already treated in the earlier chapters.

128. Application to Graduation of a Mortality Table.— As an illustration of the theory to a practical problem we present the following frequency distribution by 5-year age intervals of the number of deaths (or Σd_x by quinquennial grouping) in the recently published American-Canadian Mortality of Healthy Males, based on a radix of 100,000 entrants at age 15.

FREQUENCY DISTRIBUTION OF DEATHS BY ATTAINED AGES IN AMERICAN-CANADIAN MORTALITY TABLE

Ages	Σdx	1st Component	2d Component
15– 19	1,801	120	1,681
20– 24	1,996	230	1,766
25– 29	2,089	440	1,649
30– 34	2,120	790	1,330
35– 39	2,341	1,370	971
40– 44	2,911	2,270	641
45– 49	3,937	3,570	367
50– 54	5,527	5,400	127
55– 59	7,723	7,722	1
60– 64	10,383	10,383	
65– 69	12,987	12,987	
70– 74	14,535	14,535	
75– 79	13,807	13,807	
80– 84	10,328	10,328	
85– 89	5,464	5,464	
90– 94	1,757	1,757	
95– 99	278	278	
100–104	16	16	
	100,000	91,467	8,533

The curve represented by the d_x column is evidently a composite frequency function compounded of several series. From a purely mathematical point of view the compound curve may be considered as being generated in an infinite number of ways as the summation of separate component frequency curves. From the point of view of a practical graduation it is, however, easy to break this compound death curve up into two separate components. A mere glance at the d_x curve itself suggests a major skew frequency curve with a maximum point somewhere in the age interval from 70–75 and minor curve (practically one-sided) for the younger ages.

Let us therefore break the Σd_x column up into the two so far perfectly arbitrary parts as shown in the above table and then try to fit those two distributions to logarithmically transformed A curves.

Starting with the first component the straightforward computation of the semi-invariants is given in the table below with the provisional mean chosen at age 67.

FREQUENCY DISTRIBUTION OF DEATHS IN AMERIEAN MORTALITY TABLE
FIRST COMPONENT

Ages	x	$F(x)$	$xF(x)$	$x^2F(x)$	$x^3F(x)$
104–100	− 7	16	112	784	5,488
99– 95	− 6	278	1,668	10,008	60,048
94– 90	− 5	1,757	8,785	43,925	219,625
89– 85	− 4	5,464	21,856	87,424	349,696
84– 80	− 3	10,328	30,984	92,952	278,856
79– 75	− 2	13,807	27,614	55,228	110,456
74– 70	− 1	14,535	14,535	14,535	14,535
69– 65	− 0	12,987	0	0	0
	Σ	59,172	105,554	304,856	1,038,704
64– 60	+ 1	10,383	10,383	10,383	10,383
59– 55	+ 2	7,723	15,446	30,892	61,784
54– 50	+ 3	5,400	16,200	48,600	145,800
49– 45	+ 4	3,570	14,280	57,120	228,480
44– 40	+ 5	2,270	11,350	56,750	283,750
39– 35	+ 6	1,370	8,220	49,320	295,920
34– 30	+ 7	790	5,530	38,710	270,970
29– 25	+ 8	440	3,520	28,160	225,280
24– 20	+ 9	230	2,070	18,630	167,670
19– 15	+10	120	1,200	12,000	120,000
	Σ	32,296	88,199	350,565	1,810,037
	s_r	91,468	−17,355	655,421	771,333

Computing the semi-invariants by means of the usual formulas in paragraph 104, we have:

$\lambda_1 = -17355:91468 = -0.18974$, or mean at age $67+5(0.19)$ or at age 67.95

$\lambda_2 = 655421:91468 - \lambda_1^2 = 7.1296$

$\lambda_3 = 771333:91468 - 3\lambda_1\lambda_2 + 2\lambda_1^3 = 12.4981$

In order to determine the mathematical zero or the origin we have to solve the following cubic:

$$\lambda_2 \gamma_l^3 - 3\lambda_2^2\gamma_l^2 = \lambda_2^3, \text{ or}$$

$$12.489\gamma_l^3 - 152.511\gamma_l^2 = 362.47$$

the positive root of which is equal to 12.39. The zero point is therefore found to be situated 12.39 5-year units from the mean or at age $67.95+5(12.39)$, i. e. very nearly at age 130,

which we henceforth shall select as the origin of the co-ordinate system of the first component. We have furthermore

$$12.39 = e^{m+1.5n^2}, \text{ and } 7.1296 = e^{2m+3n^2}(e^{n^2}-1) = (12.39)^2(e^{n^2}-1),$$

the solution of which gives $n^2 = 0.04436$, $n = 0.2106$, $m = 2.4504$, all on the basis of a 5-year interval as unit. If we wish to change to a single calendar year unit we must add the natural logarithm of 5, or 1.6094, to the above value of m, which gives us $m = 4.0598$, while n remains the same. The above computations furnish us with the necessary material for the logarithmic transformation of the variate x which now may be written as

$$z = [\log(130-x) - 4.0598]:0.2106,$$

where x is the original variate or the age at death.

Having thus accomplished the logarithmic transformation we may henceforth write the generating function as

$$\Phi_0(x) = \frac{1}{.2106\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\log(130-x)-4.0598}{0.2106}\right]^2} = \varphi_0(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2:2}$$

We express now $F(x)$ by the following equation.

$$F(x) = k_0\Phi_0(x) + k_3\Phi_3(x) + k_4\Phi_4(x) + \ldots.$$

or in terms of the transformed z:

$$\varphi(z) = k_0\varphi_0(z) + k_3\varphi_3(z) + k_4\varphi_4(z) + \ldots.,$$

and proceed to determine the numerical values of k by the method of least squares.

129. Formation of Observation Equations.—The values of $\varphi_0(z)$ and its 3rd and 4th derivatives may be written down directly from the tables of Jørgensen or Charlier for various values of z as shown in detail in the following scheme on the following pages.*

(1) Age	(2) z	(3) $\varphi_0(z)$	(4) $\varphi_3(z)$	(5) $\varphi_4(z)$	(6) $k_0(3)$	(7) $k_3(4)$	(8) $k_4(5)$	(9) $F_1(z)$
15	+3.257	+.0020	−.0491	+.1029	+ 14	+ 10	− 1	23
6	3.213	.0023	.0537	.1084	17	10	1	26
7	3.170	.0026	.0586	.1146	19	12	1	30
8	3.127	.0030	.0637	.1208	22	14	1	35

*The values of z, co-ordinate with those of x are computed for all integral values of x from 15 to 100 in accordance with the previously established relation, viz. $z = [\log(130-x) - 4.0698]:0.2106$. All logarithms on base e.

(1) Age	(2) z	(3) $\varphi_0(z)$	(4) $\varphi_3(z)$	(5) $\varphi_4(z)$	(6) $k_0(3)$	(7) $k_3(4)$	(8) $k_4(5)$	(9) $F_1(z)$
9	3.085	.0034	.0688	.1249	25	15	1	39
20	3.041	.0039	.0744	.1290	29	16	1	44
1	2.999	.0044	.0803	.1331	32	17	1	48
2	2.955	.0051	.0850	.1361	38	18	1	57
3	2.911	.0057	.0919	.1382	42	19	1	60
4	2.866	.0065	.0981	.1391	48	21	1	68
25	2.821	.0074	.1044	.1390	54	22	1	75
6	2.776	.0085	.1104	.1367	63	24	1	86
7	2.730	.0096	.1168	.1328	71	25	1	86
8	2.683	.0110	.1229	.1264	81	26	1	106
9	2.637	.0123	.1286	.1159	91	27	1	117
30	2.587	.0140	.1340	.1072	103	28	1	130
1	2.542	.0150	.1387	.0943	116	29	1	144
2	2.494	.0178	.1420	.0763	131	30	1	160
3	2.445	.0201	.1462	.0576	149	31	1	179
4	2.396	.0226	.1486	.0340	166	32	0	198
35	2.346	.0255	.1496	+.0039	188	32	− 0	220
6	2.296	.0286	.1489	−.0275	210	32	+ 0	242
7	2.245	.0320	.1464	.0622	236	31	1	268
8	2.193	.0360	.1423	.0983	265	30	1	296
9	2.142	.0402	.1393	.1399	296	29	1	326
40	2.089	.0450	.1281	.1864	331	27	2	260
1	2.036	.0502	.1170	.2355	369	25	2	396
2	1.982	.0559	.1030	.2875	411	22	3	436
3	1.928	.0621	.0859	.3412	452	18	3	478
4	1.873	.0690	.0656	.3965	508	14	4	526
45	1.822	.0757	.0442	.4474	557	9	4	570
6	1.762	.0845	−.0156	.5060	622	+ 3	5	630
7	1.704	.0934	+.0154	.5596	687	− 3	6	690
8	1.647	.1028	.0487	.6082	758	10	6	754
9	1.589	.1129	.0853	.6419	832	18	6	820
50	1.529	.1239	.1255	.6893	913	27	7	893
1	1.471	.1352	.1599	.7132	994	34	7	967
2	1.409	.1479	.2114	.7349	1,089	45	7	1,051
3	1.348	.1609	.2565	.7430	1,185	54	7	1,138
4	1.286	.1745	.3022	.7307	1,288	63	7	1,231
55	1.224	.1886	.3467	.7062	1,391	74	7	1,324
6	1.160	.2035	.3907	.6642	1,501	83	7	1,425
7	1.095	.2190	.4320	.6037	1,612	92	6	1,526
8	1.030	.2347	.4688	.5257	1,730	99	5	1,636
9	0.963	.2509	.5008	.4180	1,847	106	4	1,745
60	0.896	.2672	.5257	.2911	1,965	112	3	1,856
1	0.828	.2832	.5426	.1831	2,083	115	2	1,970
2	0.758	.2994	.5489	−.0350	2,201	116	+ 0	2,085
3	0.689	.3146	.5474	+.1187	2,318	116	− 1	2,201
4	0.617	.3298	.5329	.2839	2,428	113	3	2,312

(1) Age	(2) z	(3) $\varphi_0(z)$	(4) $\varphi_3(z)$	(5) $\varphi_4(z)$	(6) $k_0(3)$	(7) $k_3(4)$	(8) $k_4(5)$	(9) $F_1(z)$
65	0.543	.3443	.5056	.4537	2,532	107	5	2,420
6	0.470	.3572	.4666	.6156	2,627	99	6	2,522
7	0.396	.3689	.4152	.7686	2,716	88	8	2,620
8	0.319	.3792	.3505	.9098	2,789	74	9	2,706
9	0.243	.3873	.2768	1.0262	2,848	59	10	2,779
70	0.164	.3937	.1918	1.1176	2,900	41	11	2,848
1	.084	.3975	.0999	1.1757	2,929	21	12	2,896
2	+0.000	.3989	+.0119	1.1968	2,937	− 3	12	2,922
3	−0.080	.3976	−.0952	1.1777	2,929	+ 20	12	2,937
4	0.164	.3937	.1918	1.1176	2,900	41	11	2,930
75	0.249	.3868	.2829	1.0180	2,848	60	10	2,898
6	0.348	.3755	.3762	.8592	2,767	80	9	2,848
7	0.425	.3645	.4368	.7043	2,686	93	7	2,772
8	0.516	.3493	.4912	.5146	2,569	104	5	2,668
9	0.608	.3316	.5303	.3069	2,444	112	3	2,553
80	0.702	.3118	.5502	+.0892	2,296	117	− 1	2,412
1	0.798	.2902	.5473	−.1204	2,134	116	+ 1	2,251
2	0.896	.2672	.5257	.3130	1,965	112	3	2,080
3	0.996	.2436	.4859	.4380	1,788	103	4	1,895
4	1.098	.2185	.4302	.5899	1,612	91	9	1,709
85	1.203	.1934	.3614	.6943	1,420	77	7	1,504
6	1.309	.1694	.2854	.7358	1,244	60	7	1,311
7	1.418	.1460	.2048	.7340	1,075	43	7	1,122
8	1.529	.1240	.1255	.6893	913	27	7	947
9	1.644	.1034	−.0505	.6106	758	+ 11	6	775
90	1.762	.0845	+.0156	.5060	622	− 3	5	624
1	1.882	.0679	.0693	.3874	500	15	4	489
2	2.004	.0536	.1090	.2663	394	23	3	374
3	2.132	.0397	.1380	.1485	292	29	1	264
4	2.260	.0310	.1478	−.0483	228	31	+ 0	197
95	2.393	.0227	.1477	+.0325	167	31	− 0	135
6	2.530	.0163	.1399	.0905	120	30	1	89
7	2.673	.0100	.1207	.1295	76	26	1	49
8	2.821	.0074	.1044	.1386	54	22	1	29
9	2.968	.0050	.0842	.1353	37	18	1	18
100	−3.124	.0028	.0640	.1203	21	14	1	6

Since the original observations of d_x are given in 5-year age intervals it becomes necessary to sum the numerical values of $\varphi_0(z)$ and its derivatives by quinquennial age groupings so as to form the required observation equations. We find thus for instance in the age interval 55-59 the following observation equation (the summation to take place from $x=55$ to $x=59$)

$$k_0\sum\varphi_0(z)+k_3\sum\varphi_3(z)+k_4\sum\varphi_4(z)=\sum\varphi(z)=o_x, \text{ or}$$

$$1.0967k_0+2.1390k_3-2.9178k_4=7722$$

Similar equations are formed for the other age intervals, resulting in the following tabular representation of the coefficients to the various k's and the observed values of the frequency distribution, $\varphi(z)$

TABLE I

TABULAR ARRANGEMENTS OF NUMERICAL DATA IN THE OBSERVATION EQUATIONS

Ages	φ_0	φ_3	φ_4	o
15–19	.0130	− .2930	+ .5730	120
20–24	.0256	− .4305	+ .5758	230
25–29	.0488	− .5833	+ .6508	440
30–34	.0903	− .7104	+ .3694	790
35–39	.1623	− .7265	− .3240	1,370
40–44	.2822	− .4996	−1.4471	2,270
45–49	.4673	+ .0896	−2.7631	3,570
50–54	.7424	+1.0555	−3.6111	5,400
55–59	1.0967	+2.1390	−2.9178	7,722
60–64	1.4942	+2.6975	− .1066	10,383
65–69	1.8369	+2.0147	+3.7739	12,987
70–74	1.9814	+ .0166	+5.7854	14,535
75–79	1.8077	−2.1174	+3.6030	13,807
80–84	1.3307	−2.5393	−1.3721	10,328
85–89	.7362	−1.0276	−3.4640	5,464
90–94	.2729	.4571	−1.3925	1,757
95–99	.0609	.5714	.6125	278
100–	.0068	.1714	.3890	16

From the above table we notice that we have 18 observation equations from which to determine the three unknown parameters k_0, k_3 and k_4. The number of equations being greater than the number of unknowns we make use of the method of least squares. While a direct application of this principle of course is feasible, it will, however, be found easier to start with an approximate solution for k_0, k_3 and k_4 and then apply the method of least squares. It will be found that in the three age intervals 65-69, 70-74 and 75-79 where the observations are most numerous the observations will be approximately satisfied by the following preliminary values of k, viz.:

$$k^1_0 = 7300,\ k^1_3 = -340 \text{ and } k^1_4 = -50.$$

Multiplying the above values of k with their respective columns in Table I, or in other words forming the products $k^1\varphi_0$, $k^1_3\varphi_3$ and $k^1_4\varphi_4$, we obtain a new table of the following form:*

*Last figure omitted.

TABLE II

Ages	a	b	c	o	Control Column s
15– 19	10	10	– 3	– 12	5
20– 24	19	15	– 3	– 23	8
25– 29	36	20	– 3	– 44	9
30– 34	66	24	– 2	– 79	9
35– 39	119	25	2	– 137	9
40– 44	206	17	7	– 227	3
45– 49	343	– 3	14	– 357	– 3
50– 54	542	–36	18	– 540	–16
55– 59	801	–73	15	– 772	–29
60– 64	1,091	–92	1	–1038	–38
65– 69	1,341	–69	–19	–1299	–46
70– 74	1,446	– 1	–29	–1454	–38
75– 70	1,320	72	–18	–1381	– 7
80– 84	971	86	7	–1033	31
85– 90	537	35	17	– 546	43
90– 94	202	–16	7	– 176	17
95– 99	45	–20	– 3	– 28	– 6
100–104	5	– 6	– 2	– 2	– 5
	9,100	–12	6	–9148	–54

The formation of the sum-products $[aa]$, $[ab]$, ... $[ao]$, $[bb]$, $[bc]$, $[bo]$ and $[cc]$, $[co]$ proceeds now in routine manner as shown in the following tables:

TABLE III

[aa]	[ab]	[ac]	[ao]	[as]
100	100	– 30	– 120	50
361	285	– 57	– 437	152
1,296	720	– 108	– 1,584	324
4,356	1,584	– 132	– 5,214	594
14,161	2,975	238	– 16,303	1,071
42,436	3,502	1,442	– 46,762	618
117,649	– 1,029	4,802	– 122,451	– 1,029
293,764	– 19,512	9,756	– 292,680	– 8,672
641,601	– 58,473	12,015	– 618,372	– 23,229
1,190,281	–100,372	1,091	–1,132,458	– 41,458
1,798,281	– 92,529	–25,479	–1,741,959	– 61,686
2,090,916	– 1,446	–41,934	–2,102,484	– 54,948
1,742,400	95,040	–23,760	–1,822,920	– 9,240
942,841	83,506	6,797	–1,003,043	30,101
288,369	18,795	9,129	– 293,202	23,091
40,804	– 3,232	1,414	– 35,552	3,434
2,025	– 900	– 135	– 1,260	– 270
25	– 30	– 10	– 10	– 25
9,211,666	– 71,016	–44,961	–9,236,811	–141,122

TABLE IV

[*bb*]	[*bc*]	[*bo*]	[*bs*]
100	− 30	− 120	50
225	− 45	− 345	120
400	− 60	− 880	180
576	− 48	− 1,896	216
625	50	− 3,425	225
289	119	− 3,859	51
9	− 42	1,071	9
1,296	− 648	19,440	576
5,329	−1,095	56,356	2,117
8,464	− 92	95,496	3,496
4,761	1,311	89,631	3,174
1	29	1,454	38
5,184	−1,296	−99,432	− 504
7,396	602	−88,838	2,666
1,225	595	−19,110	1,505
256	− 112	2,816	− 272
400	60	560	120
36	12	12	30
36,572	− 690	48,931	13,797

TABLE V

[*cc*]	[*co*]	[*cs*]
9	36	− 15
9	69	− 24
9	132	− 27
4	158	− 18
4	− 274	18
49	− 1,589	21
196	− 4,998	− 42
324	− 9,720	− 288
225	−11,580	− 435
1	− 1,038	− 38
361	24,681	874
841	42,166	1,102
324	24,858	126
49	− 7,231	217
289	− 9,282	731
49	− 1,232	119
9	14	18
4	4	10
2,756	45,244	2,349

From the above tables we may now write down the following scheme for the solution of the normal equations by means of the Gaussian algorithmus

TABLE VI

SCHEME FOR THE SOLUTION OF NORMAL EQUATIONS

921,167	− 7102	− 4496	− 923681
	55	*35*	*7122*
	3657	− 69	4893
		22	*4508*
		276	4525
	−.00771	*−.00488*	*−1.00273*
	3602	− 104	− 2229
		3	*64*
		254	16
		−.02887	*− .61882*
		251	48

Solving for the unknowns we have now

$$r_4 = 48{:}251 = .19123$$

$$r_3 = .61882 - (.19123)(-.02887) = .62434$$

$$r_0 = 1.00273 - (.62434)(-.00771) - (.19123)(-.00488) = 1.00847$$

We therefore find the following numerical values for the probable values of k_0, k_3 and k_4

$$k_0 = r_0 k'_0 = (1.00847)(7300) = 7{,}361.8$$

$$k_3 = r_3 k'_3 = (.62434)(-340) = -212.2$$

$$k_4 = r_4 k'_4 = (.19123)(-50) = -9.6$$

The next step is then to form the three columns $k_0\varphi_0(z)$, $k_3\varphi_3(z)$ and $k_4\varphi_4(z)$ for the individual ages from 15 and upwards. The formation of $\Sigma k_i\varphi_i(z)$ gives us finally the separate values by integral ages of the first component curve or $F_{\text{I}}(x)$.

If we now subtract this component from the originally observed values of the compound curve, d_x, we obtain the following values (arranged in quinquennial age groups) for the second component, $F_{\text{II}}(x)$.

Ages	$F_{II}(x)$	
0– 4	50	Hypothetical Values
5– 9	440	
10–14	1,210	
15–19	1,650	
20–24	1,719	
25–29	1,610	
30–34	1,309	
35–39	989	
40–44	715	
45–59	473	
50–54	247	
55–59	67	
	10,479	

It is of course possible to fit this particular curve type directly to a logarithmically transformed Gram or Charlier A type of frequency curves, although this will require 5 or 6 terms of the series. But still greater obstacles would be encountered if we were to attempt a graduation by means of Pearsonian curve types. The goal can, however, be reached quite readily by the introduction of a certain hypothetical device. Any reader familiar with the various types of frequency curves will readily notice that the above frequency distribution of $F_{II}(x)$ represents a truncated curve from which the curve segment corresponding to ages below 15 has been eliminated. We may therefore substitute the following, so far hypothetical values for the missing curve segment. For ages 0-4 the value of 50, for ages 5-9 the value of 440, and for the ages 10-14 the value of 1,210. The thus reconstructed histograph (shown in the above table) may now be fitted to a logarithmically transformed Gram or Charlier curve in the usual routine manner. The computations of the relative moments m_r result in the following values, for a provisional origin at the age of 17 and a unit interval of 5 years

$$m_1 = 1.8380, \; m_2 = 8.6342, \; m_3 = 39.8630$$

From which we find

$\lambda_1 = 1.838$, $\lambda_2 = 5.2560$, $\lambda_3 = 4.5824$. (Mean at age 26.19.)

The equation

$$\lambda_3 \eta^3 - 3\lambda_2^2 \eta^2 = \lambda_2^3$$

then becomes

$$4.582\eta^3 - 82.876\eta^2 = 145.200,$$

the real root of which is $\eta = 19.0$ (on basis of a 5 year unit.)

We furthermore find that $n=0.120$ and $m=2.9227+log_e 5=4.5321$, which finally brings about the transformation of the variate x by means of the formula

$$z=[\log_e(x+68.8)-4.532]:0.12$$

where x is expressed in unit intervals of 1 year.

The further determination of the coefficients k_0, k_3 and k_4 by means of the method of least squares results in the values: $k_0=947.4$, $k_3=-63.4$ and $k_4=-30.0$. Multiplying these values with their respective values of $\varphi_0(z)$, $\varphi_3(z)$ and $\varphi_4(z)$ and forming the corresponding sums we finally obtain the second component curve.

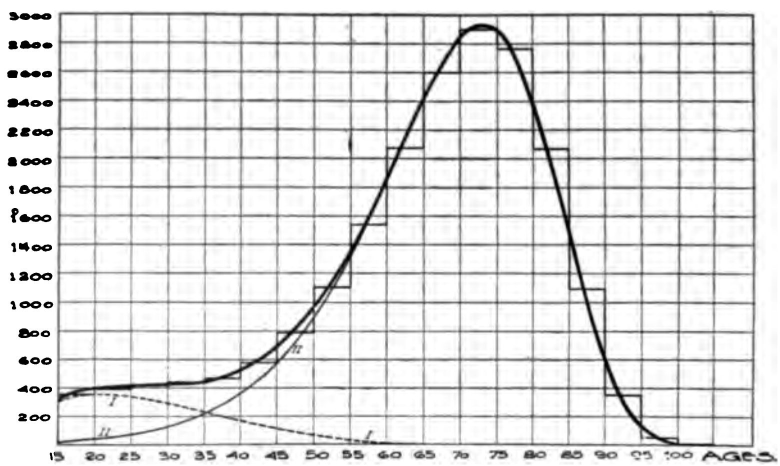


FIGURE 3

Diagram showing graduation of d_x column in the AM(5) table by a compound frequency curve of the Gram-Charlier types.

The sum of $F_I(x)$ and $F_{II}(x)$ as shown on page 255 and also in the figure gives us the final compound frequency curve or the d_x curve, from which it now is a simple matter to form the l_x or Σd_x column and its co-ordinated column of q_x.

GRADUATION OF AMERICAN MALE MORTALITY TABLE (AM(5)) BY MEANS OF A COMPOUND FREQUENCY CURVE

Age	$F_I(x)$	$F_{II}(x)$	d_x	l_x	$1000q_x$
15	21	302	323	100532	3.21
6	26	319	345	100209	3.44
7	30	332	362	99864	3.62

Age	$F_I(x)$	$F_{II}(x)$	d_x	l_x	$1000q_x$
8	35	342	377	99502	3.79
9	39	350	389	99125	3.92
20	44	354	398	98736	4.03
1	48	354	402	98338	4.11
2	57	352	409	97936	4.18
3	60	349	409	97527	4.19
4	68	343	411	97118	4.23
25	75	336	411	96707	4.25
6	86	327	412	96296	4.28
7	95	317	413	95884	4.31
8	106	308	414	95471	4.33
9	117	297	414	95057	4.36
30	130	285	415	94643	4.38
1	144	275	419	94228	4.45
2	160	261	421	93809	4.49
3	179	249	428	93388	4.58
4	198	238	436	92960	4.69
35	220	226	446	92524	4.82
6	242	213	455	92078	4.94
7	268	201	469	91633	5.12
8	296	188	484	91154	5.31
9	323	175	501	90670	5.53
40	360	164	524	90169	5.81
1	396	152	548	89645	6.11
2	436	141	577	89097	6.47
3	478	128	606	88520	6.80
4	526	117	643	87914	7.32
45	570	107	677	87271	7.76
6	630	96	726	86594	8.39
7	690	87	777	85868	9.05
8	754	78	832	85091	9.78
9	820	69	889	84259	10.55
50	893	61	954	83370	11.44
1	967	53	1020	82416	12.37
2	1051	47	1098	81396	13.49
3	1138	41	1179	80298	14.68
4	1231	36	1267	79119	16.01
55	1324	30	1354	77852	17.39
6	1425	25	1450	76498	18.95
7	1526	52	1548	75048	20.62
8	1636	18	1654	73500	22.41
9	1745	15	1760	71846	24.54
60	1856	12	1868	70086	26.65
1	1970	11	1981	68218	29.04
2	2085	9	2092	66237	31.59
3	2201	7	2208	64145	34.42
4	2312	6	2318	61937	37.41
65	2420	5	2425	59619	40.67

Age	$F_I(x)$	$F_{II}(x)$	d_x	l_x	$1000q_x$
66	2522	4	2526	57194	43.62
7	2620	3	2623	54668	47.97
8	2706	2	2708	52045	52.02
9	2779	2	2781	49337	56.37
70	2848	1	2849	46556	61.19
1	2896	0	2896	43707	66.25
2	2922	0	2922	40811	71.60
3	2937	0	2937	37889	77.51
4	2930	0	2930	34952	83.93
75	2898	0	2898	32022	90.51
6	2848	0	2848	29124	97.80
7	2772	0	2772	26276	105.48
8	2662	0	2662	23504	113.53
9	2553	0	2553	20836	122.50
80	2412	0	2412	18283	131.95
1	2251	0	2251	15871	141.84
2	2080	0	2080	13620	152.76
3	1895	0	1895	11540	164.21
4	1709	0	1709	9645	177.19
85	1504	0	1504	7936	189.52
6	1311	0	1311	6432	203.82
7	1125	0	1125	5121	219.68
8	947	0	947	3996	236.99
9	775	0	775	3049	254.18
90	624	0	624	2274	274.41
1	489	0	489	1650	296.36
2	374	0	374	1161	322.14
3	264	0	264	787	335.45
4	197	0	197	523	376.67
95	135	0	135	326	414.11
6	89	0	89	191	465.97
7	49	0	49	102	480.39
8	29	0	29	53	547.16
9	18	0	18	24	780.00
100	6	0	6	6	1000.00

It will be of interest to compare these latter values with the original values of q_x as derived by Mr. Henderson's graduation. Such a comparison is shown in the appended table for quinquennial ages.

Ages	Henderson's q_x	Fisher's q_x
15	3.46	3.21
20	3.92	4.03
25	4.31	4.25
30	4.46	4.38
35	4.78	4.82

Ages	Henderson's q_x	Fisher's q_x
40	5.84	5.81
45	7.94	7.76
50	11.58	11.44
55	17.47	17.39
60	26.68	26.65
65	40.66	40.67
70	61.47	61.19
75	91.94	90.51
80	135.74	131.95
85	197.07	189.52
90	280.35	274.41
95	387.76	414.11
100	562.50	1000.00

I think that every unbiased critic will admit that there exists a satisfactory agreement between the two tables in spite of the fact that we have worked throughout with basic data in 5-year age groups. Moreover, the actual arithmetical work in the case of the graduation by means of compounded Gram or Charlier curves is much simpler than the usual methods of graduation by Makeham's formula and mechanical interpolation formulas as employed by Mr. Henderson.* Another point speaking in favor of the frequency curve graduation is that our resulting functions are continuous functions for which standard tables of definite integrals have been prepared. It is therefore possible to use the elegant and continuous method originally introduced by Mr. Woolhouse in the computation of premiums and policy values. Unfortunately this is not the place to treat this interesting phase of the question, although we may in passing it mention that a graduation of the kind as here presented in practical computations of policy values and premiums is even easier to work with than the renowned graduation formula by Makeham, especially in the case of life contingencies involving 2 or more lives.

130. Additional Examples.—As another illustration I present the following frequency distribution (arranged in groups of 3-year intervals) of the ages of a group of 19,274 male employees of the Bell System of the American Telephone and

*I do not wish to imply these remarks as a criticism of the able graduation by Henderson, however.

Telegraph Company, which most kindly has been furnished to me through the courtesy of this company.

AGE DISTRIBUTION OF MALE EMPLOYEES IN THE BELL SYSTEM

Ages	x	$F(x)$	Ages	x	$F(x)$
13–15	0	1	46–48	11	380
16–18	1	9	49–51	12	272
19–21	2	745	52–54	13	186
22–24	3	2,264	55–57	14	141
25–27	4	3,828	58–60	15	110
28–30	5	3,801	61–63	16	72
31–33	6	2,711	64–66	17	43
34–36	7	1,918	67–69	18	17
37–39	8	1,339	70–72	19	14
40–42	9	884	73–75	20	3
43–45	10	533	76–78	21	2

Choosing the provisional lower limit at age 14 we find the following values for the crude moments or power sums s.

$$s_0 = 19{,}274,\ s_1 = 112{,}363,\ s_2 = 794{,}771,\ s_3 = 6{,}790{,}761$$

The values of the semi-invariants are

$$\lambda_1 = 5.830,\ \lambda_2 = 7.2478,\ \lambda_3 = 27.4191.$$

The resulting cubic expansion is therefore

$$27.419r^3 - 157.592r^2 = 380.731$$

for which the solution is $r = 6.1185$.

We have furthermore

$$6.1185 = e^{m+1.5n^2}$$

$$7.2478 = e^{2m+3n^2}(e^{n^2} - 1), \text{ or}$$

$$n^2 = 0.1768,\ n = 0.4205 \text{ and } m = 1.5462$$

On the basis of an interval of one year we have therefore:

$$z = [\log(x) - 13.1) - 2.645] : 0.421^*$$

as the value of the variate in the generating function $\varphi_0(z)$.

* We have $m = 1.5462 + \log 3 = 2.645$.

The values of k_0, k_3 and k_4 as determined by the method of least squares are $k_0=3064.4$, $k_3=45.1$, $k_4=80.5$, on basis of one year interval.

A comparison between the calculated and observed values (the latter being shown by single ages) is given in the attached diagram, which evidently is satisfactory for all practical purposes. I wish here to mention that an attempt by some of the

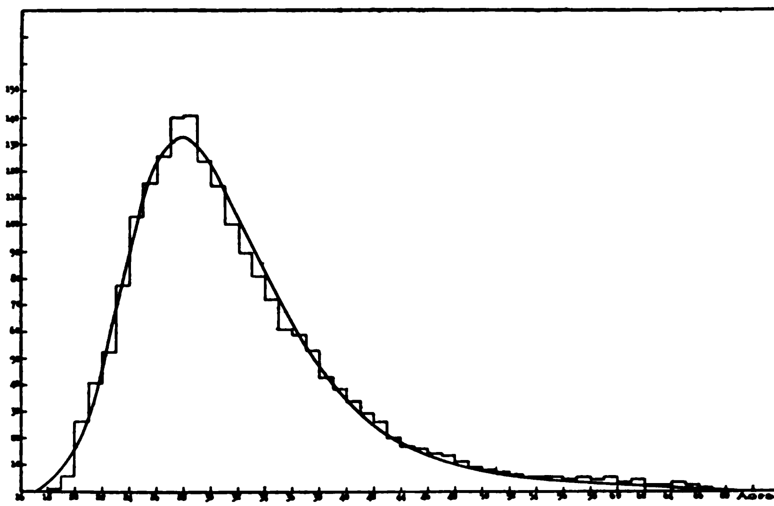


FIGURE 4

Diagram showing comparison between observed and theoretical frequency distribution of active group of male employees of the Bell System.

statistical assistants of the A. T. & T. Co. to fit the above data by means of the Pearsonian curves proved futile. Personally I have not as yet made an attempt to verify this negative result.

As a final illustration we quote from Jørgensen's monograph an application of the logarithmic transformaion of the previously discussed observations on the number of petal flowers in *Ranunculus Bulbosus*. Since the variate in this instance is integral, the observations themselves clearly indicate that there must be a lower limit, or biological zero so to speak, at 4 petal flowers.

The crude moments are then

$S_0=222$, $S_1=362$, $S_2=794$, from which we obtain

$$m = 0.0440$$
$$n = 0.5445$$
$$k_0 = 183.2,$$

so that the formula reads

$$F(x) = \frac{183.2}{.5445\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\log(x-4)-.044}{0.5445}\right]^2}$$

The detailed calculations according to this formula are shown below:

(1) log nat $(x-4)$	(2) $\log(x-4)-m$	(3) (2):n	(4) $\varphi_0(x)$	(5) $F(x)$
.0000	− .0440	− .0810	.3989	131.8
.6932	+ .6492	+1.1926	.1965	66.1
1.0986	1.0546	1.9373	.0612	20.6
1.3863	1.3423	2.4658	.0191	6.4
1.6094	1.5654	2.8756	.0064	2.2
1.7918	1.7478	3.2107	.0024	0.8

A closer fit could of course be had by adding additional terms to the series, but even with one term the agreement between calculated and observed values is quite satisfactory for all practical purposes.

CHAPTER XVII

Frequency Curves and Their Relation to the Bernoullian Series

131. The Bernoullian Series.—In Chapter IX it was shown that the general term

$$\varphi(x) = \binom{s}{x} p^x q^{s-x}$$

in the point binomial $(p+q)^s$, where p is the *a priori* probability for the happening of an event E in a single trial, represents the probability that E will happen x times and the complementary event, $\bar{E}$, $s-x$ times. We also found that the maximum term in the Bernoullian expansion of the point binomial could be written as:

$$T_m = \frac{1}{\sqrt{2\pi spq}}.$$

when s is a large number.

We wish now also to find a more simple expression for the general term, $\varphi(x)$, instead of the laborious expression involving factorials of high order.

It is evident that $\varphi(x)$ represents a frequency function of an integral variate x which can assume all positive integral values from 0 to s, and which satisfies the property of all frequency curves that

$$\sum \varphi(x) = \sum \binom{s}{x} p^x q^{s-x} = (p+q)^s = 1$$

We may therefore write $\varphi(x)$ in the form of a Gram-Charlier frequency series as

$$\varphi(x) = \Sigma c_i \varphi_i(x) \text{ for } i = 0, 3, 4, 5 \ . \ . \ .$$

This involves the computation of the semi-invariants $\lambda_r(x)$ for $r = 0, 3, 4, 5$ . . . . .

By the definition of the semi-invariants we have:

$$s_0 e^{\frac{\lambda_1\omega}{1!} + \frac{\lambda_2\omega^2}{2!} + \frac{\lambda_3\omega^3}{3!} + \ . \ . \ .} \qquad \Sigma\varphi(x) = \Sigma\varphi(x)e^{x\omega},$$

where $x=0, 1, 2, 3 \ldots\ldots s$ and $\Sigma\varphi(x)=(p+q)^s=1$, or

$$s_0 e^{\frac{\lambda_1\omega}{1!}+\frac{\lambda_2\omega^2}{2!}+\frac{\lambda_3\omega^3}{3!}+\ldots} = \varphi(0)+\varphi(1)e^{\omega}+\varphi(2)e^{2\omega}+\ldots+\varphi(s)e^{s\omega}$$

$$=\Sigma\varphi(x)e^{x\omega}=(pe^{\omega}+q)^s,$$

which for $\omega=0$ reduces to $s_0=(p+q)^s=1$.

Taking the logarithm on both sides of the above equation we have

$$f(\omega)=\frac{\lambda_1\omega}{1!}+\frac{\lambda_2\omega^2}{2!}+\frac{\lambda_3\omega^3}{3!}+\ldots=s\log(pe^{\omega}+q)$$

Now it is easily seen that

$$D^1_\omega f(\omega)=\frac{spe^{\omega}}{(pe^{\omega}+q)} \text{ or } D^1_\omega f(\omega)\,(pe^{\omega}+q)=spe^{\omega}$$

from which we find

$$D^1_\omega f(\omega)q=pe^{\omega}[s-D^1_\omega f(\omega)]$$
$$D^2_\omega f(\omega)q=pe^{\omega}[s-D^1_\omega f(\omega)-D^2_\omega f(\omega)]$$
$$D^3_\omega f(\omega)q=pe^{\omega}[s-D^1_\omega f(\omega)-2\,D^2_\omega f(\omega)-D^3_\omega f(\omega)]$$
$$D^4_\omega f(\omega)q=pe^{\omega}[s-D^1_\omega f(\omega)-3\,D^2_\omega f(\omega)-3\,D^3_\omega f(\omega)-D^4_\omega f(\omega)]$$

. . . .

. . . .

where

$$D^n_\omega f(\omega)=\frac{d^n}{d_\omega}\left[\frac{\lambda_1\omega}{1!}+\frac{\lambda_2\omega^2}{2!}+\frac{\lambda_3\omega^3}{3!}+\ldots\right]$$

Letting $\omega=0$ we have therefore successively

$$\lambda_1 q=p(s-\lambda_1)$$
$$\lambda_2 q=p(s-\lambda_1-\lambda_2)$$
$$\lambda_3 q=p(s-\lambda_1-2\lambda_2-\lambda_3)$$
$$\lambda_4 q=p(s-\lambda_1-3\lambda_2-3\lambda_3-\lambda_4)$$

. . . .

. . . .

or

$$\lambda_1 = sp$$
$$\lambda_2 = spq = \sigma^2$$
$$\lambda_3 = spq(q-p)$$
$$\lambda_4 = spq(1-6\,pq)$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

The generating function $\varphi_0(x)$ in the Gram-Charlier series may hence be written as

$$\varphi_0(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\lambda_1)^2:2\sigma^2} = \frac{1}{\sqrt{2\pi spq}} e^{-(x-sp)^2:2spq}$$

while the coefficients c_3 and c_4 of the third and fourth derivatives of $\varphi_0(x)$ according to the formulas from paragraphs 113 and 114 take on the form

$$c_3 = \frac{(-1)^3}{\sigma^3 3!} spq(q-p) = -(\sqrt{spq})^{-1}(q-p):3!$$

$$c_4 = \frac{(-1)^4}{\sigma^4 4!} spq(1-6pq) = (spq)^{-1}(1-6pq):4!$$

which serve as measures for skewness and excess.

Since p and q are proper fractions whose product never can exceed $\frac{1}{4}$ it is readily seen that for large values of s both c_3 and c_4 will become very small quantities unless either p or q should be so small that the product sp (or sq) itself would be small even when s is a large number, a case which we presently shall discuss in detail.

Apart from this exception the expression for the frequency function

$$F(x) = \varphi_0(x) + c_3\varphi_3(x) + c_4\varphi_4(x) + \;.\;.\;.$$

approaches therefore the normal probability curve of Laplace whenever s is a large number.

When, on the other hand, s in the point binomial $(p+q)^s$ is *not* a large number both c_3 and c_4 play an important role as the necessary correction factors. (For $p=q=\frac{1}{2}$ all the semi-invariants of uneven order vanish.)

The normal form of the point binomial, or $\varphi_0(x)$, was already established by Laplace who also worked out a more accurate expression for skew binomials, which expression can be shown to represent the two terms $\varphi_0(x)$ and $c_3\varphi_3(x)$

As an illustration of the above formulas we shall now try to express a few Bernoullian point binomials by means of a Gram-Charlier series.

Let us for instance try to express $(.05+.95)^{100}$ by a Gram-Charlier series. We have in this case the following values for the parameters

$$\lambda_1 = 5.0, \ \sqrt{\lambda_2} = \sigma = 2.1795, \ c_3 = -0.0688, \ c_4 = 0.00625$$

The substitution of these values in the Gram-Charlier series results in the following relative frequency distribution:

x	$F(x)$	x	$F(x)$
0	.0084	8	.0614
1	.0312	9	.0343
2	.0763	10	.0179
3	.1356	11	.0081
4	.1812	12	.0031
5	.1865	13	.0009
6	.1522	14	.0003
7	.1028	15	.0001

A similar calculation in the case of the Bernoullian binomial $(0.1+0.9)^{100}$ gives

$$\lambda_1 = 10, \ \sigma = 3, \ c_3 = -.0445, \ c_4 = 0.0021$$

with the following distribution:

x	$F(x)$	x	$F(x)$
0	.0000	12	.0984
1	.0004	13	.0732
2	.0020	14	.0502
3	.0065	15	.0322
4	.0162	16	.0194
5	.0333	17	.0109
6	.0581	18	.0058
7	.0875	19	.0027
8	.1145	20	.0012
9	.1318	21	.0005
10	.1338	22	.0002
11	.1211	23	.0001

We shall presently have occasion to compare these distributions with those obtained from a direct expansion of the point binomial.

132. Poisson's Exponential. The Law of Small Numbers.—In certain statistical series it frequently happens that the semi-invariants of higher order than zero all are equal, or that

$$\lambda_1=\lambda_2=\lambda_3= \quad . \quad . \quad . \quad . \quad =\lambda_r=\lambda.$$

We shall for the present limit our discussion to homograde series where the variate is always positive and integral, and where therefore the definition of the semi-invariants is of the form:

$$e^{\frac{\lambda\omega}{1!}+\frac{\lambda\omega^2}{2!}+\frac{\lambda\omega^3}{3!}+\ \cdots}\Sigma\varphi(x)=\Sigma\varphi(x)e^{x\omega}$$
$$=\varphi(0)e^{0\omega}+\varphi(1)e^{1\omega}+\varphi(2)e^{2\omega}+\varphi(3)e^{3\omega}+\ \ldots\ \text{, or}$$
$$e^{\frac{\lambda\omega}{1!}+\frac{\lambda\omega^2}{2!}+\frac{\lambda\omega^3}{3!}+\ \cdots}=e^{-\lambda}e^{\lambda e^{\omega}}=\Sigma\varphi(x)e^{x\omega}\text{ for } x=0,1,2,3\ \ldots\ ,$$

which also can be written as

$$e^{-\lambda}\left(1+\frac{\lambda e^{\omega}}{1!}+\frac{\lambda^2e^{2\omega}}{2!}+\ \ldots\ .\right)=\varphi(0)1+\varphi(1)e^{\omega}+\varphi(2)e^{2\omega}+\ \ldots\ .$$

The coefficient of $e^{r\omega}$ gives the relative frequency or the probability for the occurrence of $x=r$, and we henceforth find that

$$\varphi(x)=\psi(r)=\frac{e^{-\lambda}\lambda^r}{r!}$$

This is the famous Poisson Exponential, so called after the French mathematician, Poisson, who first derived this expression in his *Recherches sur la Probabilite des jugesments,* but in an entirely different manner than the one we have indicated above.

The Poisson Exponential opens now a way to the treatment of the point binomial in the exceptional cases where the product sp (or sq) is small even when s is a very large number, or when more strictly speaking the expression

$$\lim_{s=\infty} sp=\lambda$$

where λ is a finite number.

Under such conditions p (or q) must approach zero and its complementary probability q (or p) must approach unity as their limiting values.

The expressions for the semi-invariants as given in paragraph 131, *i. e.*

$$\lambda_1 = sp$$
$$\lambda_2 = spq$$
$$\lambda_3 = spq(q-p)$$
$$\lambda_4 = spq(1-6pq)$$
$$\cdots\cdots$$

will under these conditions all approach the limit sp, and the general term in the Bernoullian expansion of the point binomial can therefore be expressed by means of the Poisson exponential.

In all cases where the semi-invariants of various orders happen to be equal, or very nearly equal, the formula by Poisson will be preferable in place of the more general expansion by the Gram-Charlier series.

As an illustration we may select the simple binomial $(.001+.999)^{100}$ where the semi-invariants have the following values:

$$\lambda_1 = 0.1,\ \lambda_2 = 0.0999,\ \lambda_3 = 0.099702,\ \lambda_4 = 0.0994006,$$

and therefore may be considered as being nearly equal.

The general term, $\varphi(x)$, in this particular point binomial can therefore be written as a Poisson exponential of the form:

$$\varphi(x) = \psi(r) = e^{-0.1}0.1^r : r! \text{ for } r = 0, 1, 2, 3 \ \ldots\ .$$

The Russian statistician, Bortkewitsch, has given in his interesting and scholarly brochure *Das Gesetz der kleinen Zahlen* (1898) a four decimal place table of the Poisson exponential $e^{-\lambda}\lambda^r : r!$ for values of λ from 0.1 to 10.0. The English biometrician, Soper, in 1914 published a 6 decimal place table from $\lambda = 0.1$ to $\lambda = 15.0$. This table is found in Pearson's well-known *Tables for Biometricians*. For the above mentioned Bernoullian point binomial $(0.001+0.999)^{100}$, corresponding to the Poisson exponential $e^{-0.1}0.1^r : r!$, we find from Soper's table the following values of $\psi(r)$.

r	$\psi(r)$
0	.904837
1	.090484
2	.004524
3	.000151
4	.000004

While the exponential of Poisson requires theoretically at least, that the semi-invariants must all be of the same magnitude, it will, however, often be found that this exponential will give a fair approximation to the true observed values of the frequency curve in cases where the semi-invariants $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \dots$ do not differ greatly from each other. In this connection it is of interest to compare the fits of Poisson's exponential and the Gram-Charlier series with the true values in the binomial expansion in the three examples we have given above. Through the courteous efforts of my translator and co-editor, Miss C. Dickson, the three point binomials $(0.001+0.999)^{100}$, $(0.05+0.95)^{100}$ and $(0.10+0.90)^{100}$ have been expanded directly and the results as compared with the forms of Poisson and of Gram-Charlier are shown in the following tables:

Values of $\varphi(x)$ in Various Point Binomials

TABLE I $(0.001+0.999)^{100}$

x	Binomial	Poisson
0	.9048	.9048
1	.0906	.0905
2	.0045	.0045
3	.0001	.0002
4	.0000	.0000

TABLE II $(0.05+0.95)^{100}$

x	Binomial	Gram-Charlier	Poisson
0	.0059	.0084	.0067
1	.0312	.0312	.0337
2	.0812	.0763	.0842
3	.1396	.1356	.1404
4	.1781	.1812	.1755
5	.1800	.1865	.1755
6	.1500	.1522	.1462
7	.1060	.1028	.1044
8	.0649	.0614	.0653
9	.0349	.0343	.0363
10	.0167	.0179	.0181
11	.0072	.0081	.0082
12	.0028	.0031	.0034
13	.0003	.0009	.0013
14	.0001	.0001	.0005

TABLE III $(0.1+0.9)^{100}$

s	Binomial	Gram-Charlier	Poisson
0	.0001	.0000	.0000
1	.0003	.0004	.0005
2	.0016	.0020	.0023
3	.0059	.0065	.0076
4	.0159	.0162	.0189
5	.0339	.0333	.0378
6	.0596	.0581	.0630
7	.0889	.0875	.0901
8	.1148	.1145	.1125
9	.1304	.1318	.1251
10	.1319	.1339	.1251
11	.1199	.1211	.1137
12	.0988	.0984	.0948
13	.0743	.0732	.0729
14	.0513	.0502	.0521
15	.0327	.0322	.0347
16	.0193	.0194	.0217
17	.0106	.0109	.0128
18	.0054	.0058	.0071
19	.0026	.0027	.0037
20	.0012	.0012	.0019
21	.0005	.0005	.0009
22	.0002	.0002	.0004
23	.0000	.0000	.0002

These tables need no further explanation and demonstrate the great graduating ability of the Posison function in the case of point binomials. For although the Gram-Charlier functions give better results in the last example, we must on the other hand not forget that we had to determine 4 parameters, λ_1, σ, c_3 and c_4, while Poisson's exponential requires only the determination of the single parameter λ. The great drawback of the Poisson exponential lies in the fact that it is a discrete function, which exists only for positive integral values of the variate. It therefore does not lend itself so readily to integration as the Laplacean probability function.

It is the great achievement of the eminent Swedish astronomer and statisitican, Charlier, to have been the first to attempt to find a continuous function possessing the same powerful and flexible characteristics as the Poisson exponential. Charlier has introduced as a generating function a certain curve type which is expressed by the following formula:

$$\psi(x) = \psi_{m,n}(x) = \frac{e^{-m}}{\pi} \int_0^\pi e^{m\cos\omega} \cos\,[n\sin\omega - x\omega]d\omega$$

Using the above expression as a generating function Charlier has shown that any frequency curve can be expressed by the series

$$F(x) = \psi(x) - \frac{k_1}{1!}\Delta\psi(x) + \frac{k_2}{2!}\Delta^2\psi(x) - \frac{k_3}{3!}\Delta^3\psi(x) + \ldots .$$

At the time when Charlier introduced this function in the theory of frequency curves in a little pamphlet *Weiteres uber das Fehlergesetz* he gave only a rather short sketch of the method. He and the Danish actuary, Jørgensen, have, however, of late years further developed the method so as to make it useful in practical computations. Jørgensen has in this respect done some very neat work in supplying a method for determining the constants, k, in the above series by means of semi-invariants

We intend to treat these investigations in the forthcoming second volume of this treatise. In the meantime, however, students who encounter skew frequency distributions in their work will in nearly all cases be able to overcome the practical difficulties of a graduation by means of the logarithmic transformation of the variate, treated in the earlier chapters of this book.

133. The Law of Small Numbers.—In the case of integral variates we wish, however, to call the reader's attention to certain properties of the Poisson exponential, or probability function, which have been apparently overlooked or misunderstood by several writers, especially among the English biometricians. Somehow or other the impression among those writers has been that the frequency curve or probability function of Poisson is invariably connected with the expansion of the point binomial. Thus for instance Mr. Yule in his well-known *Introduction to the Theory of Statistics* treats it as such, while Lucy Whitaker in an article entitled "On the Poisson Law of Small Numbers" in *Biometrika* for April, 1914, subjects the whole theorem to a scathing criticism. We quote the following sentence from Miss Whitaker's article: "It might be supposed, although erroneously, that the Poisson-Exponential formula was capable

of great accuracy in addition to its great simplicity. But this is to neglect the fundamental assumptions on which it is based, namely:

(1) That the data actually correspond to a binomial
(2) That in the binomial q is small and n large."

It is true that Poisson in deriving his formula started from the Bernoullian point binomial so as to meet the cases where the normal probability curve of Laplace failed to give a close approximation, that is in the case of the limiting value when either p or q becomes very small, but s is large enough so that sp or sq remain finite; and it is probably this fact which has prompted the above remarks of Miss Whitaker. But this is really to put the cart before the horse. In the case of integral variates we can, as shown in the preceeding paragraphs, derive the Poisson probability function as a general form of frequency distributions whose semi-invariants are all equal, and it is only incidental that this property is possessed by the special binomial limit in the case mentioned above. But this property is not a *general* property of the binomial any more than it is the property of the same point binomial function to result in a Laplacean normal probability curve when the exponent s is large and $p = q = \frac{1}{2}$.

Looked at wholly from the point of view of frequency functions it is, however, not necessary to resort to the binomial limit as a base for the derivation of Poisson's formula, which can be derived directly from the definition of the semi-invariants when those peculiar parameters are considered as being of equal magnitude irrespective of their order. Now the question presents itself whether the Poisson probability curve might not, like its fellow brother, the Laplacean probability curve, be used as a generating function in expanding certain types of frequency distributions in serial form. It is to this question that the discussion is devoted in the following chapter.

CHAPTER XVIII

POISSON-CHARLIER FREQUENCY CURVES FOR INTEGRAL VARIATES

134. Charlier's B Curve.—We have already seen in the previous chapters that the Gram-Charlier frequency curve could be written as

$$F(x) = \Sigma c_i \varphi_i(x) = \Sigma c_i H_i(x) \varphi_0(x) \text{ for } i = 0, 1, 2, 3, \ldots .$$

where $\varphi_0(x)$ is the generating Laplacean probability function.

The idea now immediately suggests itself to use a similar method of expansion in the case of the Poisson probability function and to employ this exponential as a generating function in the same manner as the Laplacean function. We are, however, in the present case of the Poisson exponential dealing with a generating function which so far has been defined for positive integral values only and, therefore, represents a discrete function. For this reason it will be impossible to express the series as the sum-products of the successive derivatives of the generating function and their correlated parameters c. We can, however, in the case of integral variates express the series by means of finite differences and write $F(x)$ as follows:

$$F(x) = c_0\psi(x) + c_1\Delta\psi(x) + c_2\Delta^2\psi(x) + \ldots . \qquad (I)$$

where $\psi(x) = e^{-m} m^x : x!$ for $x = 0, 1, 2, 3, \ldots .$, and

$$\Sigma\psi(x) = 1,$$

$$\Delta\psi(x) = \psi(x) - \psi(x-1),$$

$$\Delta^2\psi(x) = \Delta\psi(x) - \Delta\psi(x-1) = \psi(x) - 2\psi(x-1) + \psi(x-2).$$

The series (I) is known as the Poisson-Charlier frequency series or Charlier's B type of frequency curves.

The semi-invariants of these frequency series are given by the following relation:

$$e^{\lambda_1\omega + \lambda_2\frac{\omega^2}{2!} + \lambda_3\frac{\omega^3}{3!} + \ldots} = \sum_{x=0}^{x=\infty} [c_0\psi(x) + c_1\Delta\psi(x) + c_2\Delta^2\psi(x) + \ldots . .]\, e^{x\omega}$$

Expanding and equating the co-efficients of equal powers of ω we have:

$$
\begin{aligned}
&\lambda_0 = 1 = c_0\Sigma\psi(x) \text{ or } c_0 = 1 \\
&\lambda_1 = \quad \Sigma x\,(\psi(x) + c_1\Delta\psi(x) + c_2\Delta^2\psi(x) + \ldots\ldots) \qquad (II)\\
&\lambda_1^2 + \lambda_2 = \Sigma x^2(\psi(x) + c_1\Delta\psi(x) + c_2\Delta^2\psi(x) + \ldots\ldots) \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}
$$

We now have

$$\Sigma\psi(x) = 1, \text{ and}$$

$$\Sigma x\psi(x) = \Sigma me^{-m}m^{x-1}:(x-1)! = m\Sigma\psi(x-1) = m.$$

We also find from well-known formulas of the calculus of finite differences that*

$$
\begin{aligned}
&\Sigma x^2\psi(x) = m^2 + m \\
&\Sigma x\Delta\psi(x) = -1 \\
&\Sigma x\Delta^2\psi(x) = 0 \\
&\Sigma x^2\Delta\psi(x) = -(2m+1) \\
&\Sigma x^2\Delta^2\psi(x) = 2
\end{aligned}
$$

Substituting these values in (II) we obtain

$$
\begin{aligned}
&\lambda_1 = m - c_1 \\
&\lambda_1^2 + \lambda_2 = m^2 + m - (2m+1)c_1 + 2c_2
\end{aligned}
$$

By letting $m = \lambda_1$ we can make the coefficient c_1 vanish, which results in

$$
\begin{aligned}
&\lambda_1 = m \\
&c_2 = \tfrac{1}{2}[\lambda_2 - m]
\end{aligned}
$$

where the two semi-invariants λ_1 and λ_2 are calculated around the natural zero of the number scale as origin.

For the above discussion we have limited ourselves to the determination of the three constants m, c_0 and c_2. It is easy, however, to find the higher parameters $c_3, c_4, c_5, \ldots\ldots$ from the relations between the moments of the Poisson function and the semi-invariants of order 3, 4, 5, ... etc. Charlier usually

*These formulas can also be derived from the definition of the semi-invariants and the well-known relations between moments and semi-invariants as given on page 237, when we remember that according to our definition all semi-invariants in the Poisson exponential are equal to m.

calls the parameter m the *modulus* and c_2 the *eccentricity* of the B curve.

135. Numerical Examples.—As an illustration of the application of the Poisson-Charlier series we select the previously mentioned series of observations on alpha particles radiated from a bar of Polonium as determined by Rutherford and Geiger and shown on page 174 of this treatise. We are here dealing with integral variates which can assume positive values only and the observations are therefore eminently adaptable to the treatment by Poisson-Charlier curves. Selecting the natural zero as the origin of the co-ordinate system we find that the first two semi-invariants are of the form

$$\lambda_1 = 3.8754,\ \lambda_2 = 3.6257, \text{ and we therefore have:}$$

$$m = \lambda_1 = 3.88;\ c_2 = \tfrac{1}{2}[\lambda_2 - m] = -0.125$$

The equation for the frequency distribution of the total $N = 2608$ elements therefore becomes

$$F(x) = N[\psi_{3.88}(x) + (-0.125)\Delta^2\psi_{3.88}(x)].$$

The table below gives the values as fitted to the curve, $F(x)$:

Alpha Particles Discharged from Film of Polonium
(Rutherford and Geiger)
$N = 2608,\ m = 3.88,\ c_2 = -0.125$

(1) x	(2) $\psi(x)$	(3) $\Delta^2\psi(x)$	(4) $N\times(2)$	(5) $N\times(3)\times c_2$	(6) (4)+(5)
0	.020668	+.020668	53.9	− 6.7	47
1	.080156	+.038820	209.0	−12.7	196
2	.155455	+.015811	405.4	− 5.2	400
3	.201015	−.029739	524.2	+ 9.7	533
4	.194967	−.051608	508-5	+16.8	525
5	.151265	−.037654	394.5	+12.3	407
6	.097850	−.009714	254.9	+ 3.2	258
7	.054249	+.009814	141.2	− 3.2	138
8	.026316	+.015668	68.7	− 5.1	64
9	.011351	+.012968	29.6	− 4.2	25
10	.004407	+.008021	11.5	− 2.6	9
11	.001555	+.004092	4.1	− 1.2	3
12	.000503	+.001800	1.3	− 0.6	1
13	.000150	+.000699	0.4	− 0.2	0
14	.000042	+.000245	0.1	− 0.1	0
15	.000010	+.000076	0.0	− 0.0	0
16	.000003	+.000025	——	——	0
17	.000001	+.000005	——	——	0

Bateman has in *Philosophical Transactions* (1902) given a theoretical frequency distribution of the above series of observations wherein he develops the Poisson probability function, being ignorant of the previous demonstration by Poisson. In a later note he mentions that the formula was given by the French mathematician in his work on probabilities, published in 1837.

Bateman's calculation includes, however, only the first term of the Poisson-Charlier series and is, therefore not so close as the above fit.

As a second example we offer our old friend, the distribution of flower petals in *Ranunculus Bulbosus.* Selecting the zero point at $x=5$ and computing the semi-invariants in the usual manner we obtain the following equation for the frequency curve.

$$F(x)=222\psi(x)+31.5\Delta^2\psi(x),\ m=0.631$$

A comparison between calculated and observed values follows below:

x	$F(x)$	Obs.
5	134.9	133
6	51.6	55
7	22.5	23
8	9.5	7
9	2.9	2
10	0.6	2

136. Transformation of the Variate.—For integral variates we have shown that the Poisson frequency curve possesses the important property that all its semi-invariants are equal. Now while a frequency distribution of a certain integral variate, x, may perhaps *not* possess this property, it may, however, very well happen after a suitable linear transformation has been made, that the variate thus transformed will be subject to the laws of Poisson's function.

Let $z=ax-b$ represent the linear transformation which is subject to the above laws with a series of semi-invariants all equal to m.

These semi-invariants according to the properties set forth in paragraph (104) are therefore

$$m = \lambda_1(z) = a\lambda_1(x) - b$$
$$m = \lambda_2(z) = a^2\lambda_2(x)$$
$$m = \lambda_3(z) = a^3\lambda_3(x)$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

and our problem is to find the unknown parameters a, b and m.

Simple algebraic methods, which it will not be necessary to dwell upon, give the following results:

$$a = \lambda_2 : \lambda_3$$
$$m = \lambda_2^3 : \lambda_3^2$$
$$b = a\lambda_1 - m$$

As a numerical illustration of this transformation we choose from Jørgensen a series of observations by Davenport on the frequency distribution of glands in the right foreleg of 2000 female swine.

No. of Glands	0	1	2	3	4	5	6	7	8	9	10
Frequency . .	15	209	365	482	414	277	134	72	22	8	2

The values of the three first semi-invariants are

$$\lambda_1 = 3.501,\ \lambda_2 = 2.825,\ \lambda_3 = 2.417,$$
$$a = 2.825 : 2.417 = 1.168$$
$$m = 2.825^3 : 2.417^2 = 3.859$$
$$b = (1.168)(3.501) - 3.859 = 0.230.$$

The new variable then becomes $z = ax - b$ and the transformed Poisson probability function takes on the form:

$$\psi(z) = \frac{e^{-m}m^z}{z!}$$

In general, however, we will find that z is not a whole number and the expression $z!$ therefore has no meaning from the point of view of factorials at least. This difficulty may, however, be overcome through the introduction of the well-known Gamma Function, $\Gamma(z+1)$, which holds true for any positive

or negative real value of z and which in the case of integral values of z reduces to $\Gamma(z+1)=z!$

Hence we can write the transformed Poisson probability function as

$$\psi(z)=\frac{e^{-m}m^z}{\Gamma(z+1)}$$

Tables to 7 decimal places of the Gamma Function, or rather for the expression $-\log\,\Gamma(z+1)$, have been computed by Jørgensen in his aforementioned book from $z=-5$ to $z=15$, progressing by intervals of 0.01.

By means of this table and the tables of ordinary logarithms it is now easy to find the values of $\psi(z)$ in the case of the example relating to the number of glands in female swine. The detailed computation is shown below.*

(1)	(2)	(3)	(4)	(5)	(6)	7
x	z	$-\log\Gamma(z+1)$	$\log m^z$	$(3)+(4)+\log e^{-m}$	$\psi(z)$	$F(x)$
0	−.230	.9209	.8651	.1101–2	.0129	30.1
1	+.938	.0108	.5500	.8849–2	.0767	179.2
2	2.106	.6555	.2350	.2146–1	.1639	382.9
3	3.274	.0679	.9199	.3119–1	.2051	479.1
4	4.442	.3216	.6048	.2505–1	.1780	415.8
5	5.610	.4547	.2897	.0685–1	.1171	273.6
6	6.778	.4904	.9746	.7891–2	.0615	143.7
7	7.946	.4446	.6595	.4282–2	.0268	62.6
8	9.114	.3285	.3444	.9970–3	.0099	23.1
9	10.282	.1506	.0294	.5041–3	.0032	7.5
10	11.450	.9177	.7143	.9561–4	.0009	2.1

137. The Bernoullian Series Expressed by B Curves. In the case of the Bernoullian point binomial we have $\lambda_1=sp$ and $\lambda_2=spq$. If we now wish to express the general term, $\varphi(x)$, in the binomial by a Poisson-Charlier series we evidently have

$$\varphi(x)=\psi(x)+c_2\Delta^2\psi(x) \text{ where}$$

$$\psi(x)=e^{-m}m^x:x!$$

Now $m=\lambda_1=sp$, and $c_2=\frac{1}{2}(\lambda_2-m)=\frac{1}{2}(spq-sp)=-\frac{1}{2}sp^2$. As an illustration of this expansion we may again look at the

*The characteristics of the logarithms have been omitted in this table (except in column 5) and only the positive mantissas are shown. Column 7 represents the 2000 individual observations pro rated according to column 6.

point binomial $(0.1+0.9)^{100}$, discussed on page (268). There we have:

$$m=10,\ c_2=-\tfrac{1}{2}, \text{ while}$$

$$\varphi(x)=\psi_{10}(x)-\tfrac{1}{2}\Delta^2\psi_{10}(x).$$

The actual computation by means of this formula results in the following tabular representation.

x	Poisson-Charlier	x	Poisson-Charlier
0	.0000	12	.0986
1	.0002	13	.0744
2	.0016	14	.0514
3	.0059	15	.0330
4	.0159	16	.0197
5	.0340	17	.0108
6	.0599	18	.0055
7	.0893	19	.0025
8	.1149	20	.0011
9	.1301	21	.0005
10	.1314	22	.0001
11	.1194	23	.0000

A comparison of this series with the actual expansion of the point binomial by Miss Dickson on page (268) leaves little to be desired in the way of exactitude. The fit to the true binomial is even closer than that of Gram-Charlier series in spite of the fact that only two parameters enter into the determination of Poisson-Charlier's curves while four parameters are required for the Gram-Charlier curves.

138. Remarks on Mr. Keynes' Criticisms.—From the above discussion it is evident that the Bernoullian point binomial can always be represented without difficulty by either the Gram-Charlier or the Poisson—Charlier frequency curves. This point is of interest in connection with some wholly misleading and erroneous statements regarding Laplace's analysis of the Bernoullian Theorem by Mr. J. M. Keynes, in his recently issued "*Treatise on Probability.*

On page 358 Mr. Keynes points out the assymmetry in the Bernoullian expansion and claims that the want of symmetry is generally being overlooked, "and it is not uncommon to assume that the probability of a given divergence less than pm is equal to that of the same divergence in excess of pm, and, in general, that the probability of the frequency's exceeding pm in a set of m trials is *equal* to that of falling short of pm."

No real mathematician, and least of all Laplace, has ever claimed the presence of symmetry as being general in the case of the Bernoullian

series. Those who have fallen into that error are economists and statisticians who like Mr. Keynes are ignorant of the true assumptions underlying Bernoulli's and Laplace's demonstrations. Every mathematician knows for instance that the annual loss ratios for total permanent disability benefits in workmen's compensation assurance on the basis of a small sample payroll of say 100,000 Kroner can assume all possible real values from zero and upwards, and losses *in excess* of 100,000 are indeed possible. Statistics from certain Scandinavian industries show that the average annual loss ratio in respect to total permanent disability is about 1 Krone for each 100,000 Kroner of payroll exposure. We know, however, of a certain instance in the case of a small industrial establishment with a payroll of nearly 100,000 Kroner where the losses in a single year were more than 20,000 Kroner. In the great majority of cases, however, the annual losses are nil. We are, therefore, dealing with a decidedly skew Poisson—Charlier frequency curve.

Mr. Keynes' example is in fact less striking. He considers the case of throwing aces in 60 successive throws with a die, and he remarks that the ace cannot appear less often than not at all, whereas it may well appear more than 20 times. This, of course, is self evident and realized by Laplace in his analysis, and there is no valid reason for dwelling at length on such simple matters. But the English scholar is evidently greatly impressed by these very simple considerations for he continues in a most charming and naive manner as follows:

"The actual measurement of this want of symmetry and the determination of the conditions, in which it can be safely neglected, involves laborious mathematics of which I am only acquainted with one direct investigation, that published in the *Proceedings of the London Mathematical Society* by Mr. T. C. Simmons."

How this charming and naive statement reminds one of the playful sophistries of a bright and impish, but not necessarily bad, small boy, trying to offer some excuse and explanation for his mischievous pranks, while he at the same time is wholly unmindful of the fact that his explanations and excuses are the most damning evidence of his own guilt.

Here we have Mr. Keynes, a successful writer of economic subjects, posing as a critic of such intellectual giants in the realm of mathematical science as Bernoulli, Laplace and Poisson (which of course presupposes that he must have read very carefully the various writings of those old masters); who calmly and in the most innocent manner admits that of the "laborious" mathematics involved in this question he is only acquainted with a rather clumsy demonstration published in 1896.

While I have not the slightest doubt as to the veracity of these facts so far as Mr. Keynes is concerned, it will be of interest to see what the actual historical facts are. Now in so far as the measurement of the assymmetry or skewness of the Bernoullian point binomial is concerned this was already performed by Laplace himself, an accomplishment which in itself creates a degree of doubt in the reader's mind as to whether Mr. Keynes really has studied Laplace with the necessary care required of one who poses as a critic of the great Frenchman. Harald Westergaard, the

eminent Danish scholar, whose fame as a statistician surely rests on a far more secure foundation than that of Keynes takes in his *Statistikens Teori i Grundrids* (Copenhagen, 1915), special pains to point out that Laplace was the first to give a mathematical measure of the skewness in a Bernoullian frequency distribution.

The Danish actuary, Gram, long before the intellect of our recent English critic saw the light, derived his general series for frequency curves, which of course also applies to the Bernoullian case. Thiele in his *Almindelig Iagttagelseslaere*, at a time when the young Keynes probably was being piloted by his nursemaid or governess, discussed the same series from the point of view of semi-invariants. Later on Charlier continued in direct line from where Laplace and Poisson concluded their labours.

The necessary corrections to the generating functions, whether these be Laplace's or Poisson's probability curves, as derived by these Scandinavian writers are given in Chapters XVII and XVIII of this treatise. Not a single one of these demonstrations requires the "laborious" mathematics as mentioned by Mr. Keynes. The fact that our romancing English economist evidently is in blissful ignorance of the fundamental work by the Scandinavian school is, however, no excuse for his superficial knowledge of the expansion of statistical series, since much of this work has appeared in English.

Mr. Keynes' misconception of the real significance of the Law of Small Numbers and his criticism of Bortkewicz may possibly also be traced to his apparent ignorance of the work of the Scandinavian authors. His criticism moves practically along the same lines as that of the views held by Miss Whitaker and described on page 270 of this treatise. He, like Miss Whitaker, fails to realize that the generating Poisson probability function arises from the more general fact that all its semi-invariants are equal, rather than from the more special fact that one of the limiting values of the point binomial reduces to a Poisson frequency curve. The very fact that Mr. Keynes in his large volume never mentions the semi-invariants leads one to inquire whether those important statistical parameters remain a closed book to him.*

*Similar immature views as those held by Keynes and Miss Whitaker are also expressed by Mr. A. Mowbray in the May, 1920, *Proceedings of the Casualty Actuarial Society of America* (page 197). Judging from his tedious and laborious analysis Mowbray evidently never has heard of the semi-invariants.

z	$\varphi_0(z)$	$\varphi_3(z)$	$\varphi_4(z)$	$\varphi_5(z)$	$\varphi_6(z)$	$\int_{-\infty}^{z}\varphi_0(z)dz$
0.00	.3989	+0.0000	+1.1968	−0.0000	−5.9841	.5000
.05	.3984	.0597	1.1894	.2983	5.9319	.5199
.10	.3970	.1187	1.1671	.5915	5.7763	.5398
.15	.3945	.1762	1.1304	.8743	5.5208	.5596
.20	.3910	.2315	1.0799	1.1420	5.1711	.5793
.25	.3867	.2840	1.0165	1.3900	4.7351	.5987
.30	.3814	.3330	.9413	1.6142	4.2223	.6179
.35	.3752	.3779	.8556	1.8111	3.6439	.6368
.40	.3683	.4184	.7607	1.9777	3.0122	.6554
.45	.3605	.4539	.6583	2.1117	2.3414	.6736
.50	.3520	.4841	.5501	2.2114	1.6448	.6915
.55	.3429	.5088	.4378	2.2760	.9371	.7088
.60	.3332	.5278	.3231	2.3052	− .2324	.7257
.65	.3230	.5411	.2078	2.2995	+ .4555	.7422
.70	.3123	.5486	+ .0937	2.2601	1.1135	.7580
.75	.3011	.5505	− .0176	2.1888	1.7298	.7734
.80	.2897	.5469	.1247	2.0880	2.2938	.7881
.85	.2780	.5381	.2260	1.9604	2.7964	.8023
.90	.2661	.5245	.3203	1.8095	3.2304	.8159
0.95	.2541	.5062	.4067	1.6387	3.5898	.8289
1.00	.2420	.4839	.4839	1.4518	3.8715	.8413
1.05	.2299	.4580	.5516	1.2529	4.0735	.8531
1.10	.2179	.4290	.6091	1.0458	4.1958	.8643
1.15	.2059	.3973	.6561	.8346	4.2403	.8749
1.20	.1942	.3635	.6925	.6230	4.2103	.8849
1.25	.1826	.3282	.7185	.4147	4.1107	.8944
1.30	.1714	.2918	.7341	.2130	3.9475	.9032
1.35	.1604	.2549	.7399	− .0209	3.7278	.9115
1.40	.1497	.2180	.7364	+ .1590	3.4595	.9192
1.45	.1394	.1815	.7243	.3244	3.1510	.9265
1.50	.1295	.1457	.7042	.4735	2.8109	.9332
1.55	.1200	.1111	,6772	.6051	2.4481	.9394
1.60	.1109	.0780	.6440	.7181	2.0712	.9452
1.65	.1023	.0468	.6057	.8121	1.6886	.9505
1.70	.0940	+ .0175	.5632	.8870	1.3079	.9554
1.75	.0863	− .0094	.5173	.9431	.9363	.9599
1.80	.0790	.0341	.4692	.9809	.5801	.9641
1.85	.0720	.0563	.4195	1.0014	+ .2450	.9678
1.90	.0656	.0760	.3693	1.0058	− .0646	.9713
1.95	.0596	.0933	.3192	.9955	.3452	.9744

z	$\varphi_0(z)$	$\varphi_3(z)$	$\varphi_4(z)$	$\varphi_5(z)$	$\varphi_6(z)$	$\int_{-\infty}^{z}\varphi_0(z)dz$
2.00	.0540	−0.1080	−0.2700	+0.9718	−0.5939	.9772
2.05	.0488	.1203	.2223	.9366	.8091	.9798
2.10	.0440	.1302	.1765	.8915	.9899	.9821
2.15	.0396	.1380	.1332	.8382	1.1362	.9842
2.20	.0355	.1436	.0927	.7784	1.2488	.9861
2.25	.0317	.1473	.0554	.7139	1.3291	.9878
2.30	.0283	.1492	− .0214	.6460	1.3788	.9893
2.35	.0252	.1495	+ .0092	.5764	1.4004	.9906
2.40	.0224	.1483	.0362	.5064	1.3965	.9918
2.45	.0198	.1459	.0598	.4372	1.3701	.9929
2.50	.0175	.1424	.0800	.3697	1.3242	.9938
2.55	.0154	.1380	.0968	.3050	1.2619	.9946
2.60	.0136	.1328	.1105	.2438	1.1865	.9953
2.65	.0119	.1270	.1213	.1865	1.1007	.9960
2.70	.0104	.1207	.1293	.1338	1.0076	.9965
2.75	.0091	.1141	.1347	.0858	.9098	.9970
2.80	.0079	.1073	.1379	.0429	.8097	.9974
2.85	.0069	.1003	.1391	+ .0049	.7095	.9978
2.90	.0060	.0934	.1385	− .0281	.6110	.9981
2.95	.0051	.0865	.1364	.0563	.5159	.9984
3.00	.0044	.0798	.1330	.0798	.4255	.9987
3.05	.0038	.0732	.1284	.0989	.3407	.9989
3.10	.0033	.0669	.1231	.1140	.2624	.9990
3.15	.0028	.0609	.1171	.1253	.1911	.9992
3.20	.0024	.0552	.1106	.1332	.1271	.9993
3.25	.0020	.0499	.1039	.1381	.0705	.9994
3.30	.0017	.0449	.0969	.1404	− .0213	.9995
3.35	.0015	.0402	.0899	.1403	+ .0207	.9996
3.40	.0012	.0359	.0829	.1384	.0561	.9997
3.45	.0010	.0318	.0761	.1348	.0849	.9997
3.50	.0009	.0283	.0694	.1300	.1078	.9998
3.55	.0007	.0249	.0631	.1242	.1254	.9998
3.60	.0006	.0219	.0570	.1175	.1380	.9998
3.65	.0005	.0193	.0513	.1104	.1464	.9999
3.70	.0004	.0168	.0460	.1030	.1510	.9999
3.75	.0004	.0146	.0410	.0954	.1525	.9999
3.80	.0003	.0127	.0365	.0878	.1512	.9999
3.85	.0002	.0110	.0323	.0803	.1478	.9999
3.90	.0002	.0095	.0284	.0730	.1426	.99995
3.95	.0002	.0081	.0249	.0660	.1361	.99997

ADDENDA.

APPENDIX AND BIBLIOGRAPHICAL NOTES.

CHAPTER I.

Page 3. The establishment of the relations between hypothetical judgments and probabilities is probably first due to F. C. LANGE. See also the discussion in SIGWART'S "Logic" (English translation, Macmillan Co., New York, 1904). A defense of the "principle of insufficient reason" as opposed to the view of von Kries is given by K. STUMPF ("Über den Begriff der matematischen Wahrscheinlichkeit") *Ber. bayr. Ak.* (*phil. Kl.*), 1892. For a further discussion of the philosophical aspect the reader is advised to consult "Theorie und Methoden der Statistik" (Tubingen, 1913) by the Russian statistician, A. KAUFMANN.

CHAPTER II.

Page 21. An interesting account of the application of the theory of probabilities to whist is given by POOLE in "Philosophy of Whist Play" (New York and London, 1883). Page 23. Example 6. This is a general case of the so-called game of "Treize" or "Recontre" first discussed by MONTMORT in his "Essai sur les Jeux des Hazards" (1708). "Thirteen cards numbered 1, 2; 3, . . . up to 13 are thrown promiscuously into a bag and then drawn out singly; required the chance that once at least the number on a card shall coincide with the number expressing the order in which it is drawn." This is one of the stock problems in probability and has been discussed by nearly all the leading classical writers on the subject.

CHAPTER IV.

The close connection between probability and symbolic logic is admirably discussed by the Italian mathematician, PEANO, in various of his mathematical texts. Page 42. Example 19. See also the discussion by R. HENDERSON in "Mortality and Statistics" (New York, 1915).

CHAPTER V.

38. The moral expectation has been discussed by HARALD WESTERGAARD in "Tidsskrift for Matematik" (1878) and in "Smaaskrifter tilegnede C. F. Krieger" (Copenhagen, 1889).

CHAPTER VI.

A German translation with explanatory notes of BAYES'S brochure has recently appeared in the series of "Ostwald's Klassiker."

Page 74. The double integral in the numerator of (IV) is evidently of the form:

$$\iint_{(A)} F(y_1, y_2)dy_1dy_2,$$

where the contour of the field of integration (A) is defined by means of the relations:

$$\alpha < y_1y_2 < \beta, \quad 0 < y_1 < 1 \quad \text{and} \quad 0 < y_2 < 1.$$

The field of integration is thus the area swept out by the hyperbola $y_1y_2 = \alpha$, the straight line $y_2 = 1$, the hyperbola $y_1y_2 = \beta$ and the straight line $y_1 = 1$.

Changing the variables by means of the transformation:

$$y_1y_2 = y = \varphi(y, z) \quad \text{and} \quad 1 - y_1 = z(1 - y) = \psi(y, z)$$

we get the following new double integral

$$\iint_{(A_1)} F[\varphi(y, z), \quad \psi(y, z)]|J|dydz \qquad (|J| \text{ taken as absolute value}),$$

where J is the Jacobian or functional determinant defined by the formula:

$$J = \begin{vmatrix} \dfrac{\partial\varphi}{\partial y} & \dfrac{\partial\varphi}{\partial z} \\ \dfrac{\partial\psi}{\partial y} & \dfrac{\partial\psi}{\partial z} \end{vmatrix} = \frac{\partial\varphi}{\partial y}\frac{\partial\psi}{\partial z} - \frac{\partial\varphi}{\partial z}\frac{\partial\psi}{\partial y}.$$[1]

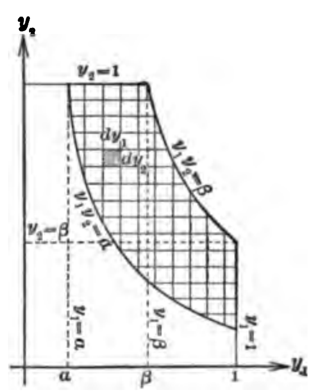
For

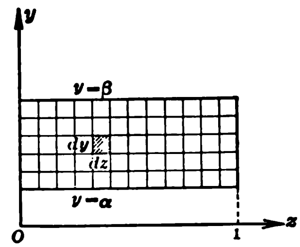
$$y_2 = \frac{y}{1 - z(1 - y)} = \varphi(y, z),$$

$$y_1 = 1 - z(1 - y) = \psi(y, z),$$

$$|J| = \left\| \begin{matrix} \dfrac{1 - z}{[1 - z(1 - y)]^2} & \dfrac{y(1 - y)}{[1 - z(1 - y)]^2} \\ z & -(1 - y) \end{matrix} \right\| = \frac{1 - y}{1 - z(1 - y)}$$

The transformation in a double integral implies in general three parts (1) the expression of $F(y_1y_2)$ in terms of y, z(2) the determination of the new system of limits (3) substitution of dy_1dy_2. The solution of the third part we just gave above. The solution of the two first is purely algebraically. The first part is a straightforward simple problem which should present no difficulty





[1] See GOURSAT: "Mathematical Analysis" (New York, 1904) pages 266–67.

whatsoever to the student and which in conjunction with (3) brings the integrands on the form given in formula (V).

The easiest way to determine the new system of limits is probably by constructing the contour in the new field of integration. The hyperbolas $y_1y_2 = \alpha$ and $y_1y_2 = \beta$ are in the new field of integration changed into the two straight lines $y = \alpha$ and $y = \beta$ which determine the limits for the variable y. A mere inspection of the expressions for $\varphi(y, z)$ and $\psi(y, z)$ shows that the two straight lines $y_2 = 1$ and $y_1 = 1$ become in the new field $z = 1$ and $z = 0$ which are the limits for z.

The contour (A_1) simply becomes a rectangle bounded by the straight lines $z = 0$, $y = \beta$, $z = 1$ and $y = \alpha$. The complete transformation finally brings the numerator on the form as given in (V).

Page 75. The question put by Mr. Bing is simply the determination of a future event by means of Bayes's Rule. The limits α and β become 0 and 1 respectively and the contour of the field of integration simply becomes the area bounded by $y_1y_2 = 0$, $y_2 = 1$, $y_1y_2 = 1$ and $y_1 = 1$, *i. e.*, the area enclosed between the two axis, the line $y_2 = 1$, the hyperbola $y_1y_2 = 1$ and the line $y_1 = 1$. The transformed contour becomes a square with side equal to unity.

CHAPTER VII.

Page 83. The criticism by the English empiricists is to a certain extent due to a misconception of the Bernoullian Theorem. "This theorem," VENN says, "is generally expressed somewhat as follows: That in the long run all events will tend to occur with a frequency proportional to their objective probabilities." Any one giving careful attention to the deduction of the famous theorem will, however, readily notice the fallacy of such a view. Not the actual absolute frequencies of the events but the *mathematical expectations of such events* are proportional to the a priori mathematical probability p. The fallacy of Mr. Venn lies in his confusing an actual event with its mathematical expectation. In other words, he makes the Bernoullian Theorem appear as a regular hypothetical judgment whereas as a matter of fact it is a simple probability judgment. If one is to take such an erroneous view of the Bernoullian Theorem one may even be reconciled with another startling statement by Venn that "If the chance (against the happening of a certain event) be 1,023 to 1 it undoubtedly will happen once in 1,024 trials."

For a clear presentment of the empirical methods and their relation to mathematical probabilities and deductive methods see v. BORTKIEWICZ "Kritische Betrachtungen zur theoretischen Statistik" (Jahrb. f. N.-Oe. u. Stat. 3 Folge, Ed. 8, 10, 11) and "Die statistischen Generalisationen" (*Scientia*, Vol. V). v. Bortkiewicz is but one of the brilliant school of Russian statisticians who has made a thorough study of the philosophical aspects of statistics. The induction method of J. S. MILL is carried much farther and put on a far sounder basis than that originally given by Mill in the brochure "Die Statistik als Wissenschaft" by A. A. TSCHUPROFF as well as in his Russian text "Researches on the Theory of Statistics." The main ideas of the Russian writers are also found in KAUFMANN's "Theorie und Methoden der Statistik" (Tübingen, 1913).

CHAPTER IX.

Page 95. For a closer approximation of $n!$ see FORSYTH, A. R., "On an Approximate Expression for $x!$" (Brit. Ass. Rep., 1883). Page 107. In this discussion it must be remembered that the variables are independent of each other. The formula: $\epsilon(k\alpha) = k\epsilon(\alpha)$ is self evident, but may be proved as follows:

$$e(k\alpha) = ksp = ke(\alpha),\ e[k\alpha - e(k\alpha)]^2 = k^2e[\alpha - e(\alpha)]^2 = k^2\epsilon^2(\alpha), \text{ or } \epsilon(k\alpha) = k\epsilon(\alpha).$$

Page 115. See also a similar discussion by WESTERGAARD in "Mortalität und Morbilität" (Jena, 1902), page 187.

CHAPTER XI.

The still unfinished series of monographs by CHARLIER are found in various volumes of *Meddelande från Lunds Astronomiska Observatorium* (Lund, Stockholm) and in *Svenska Aktuarieföreningens Tidsskrift* (Stockholm).

Page 137. Since all statistical characteristics to greater or less extent are effected with mean errors due to sampling it is of importance to be able to determine such mean errors in simple algebraic terms. We shall for the present confine ourselves to the mean and the dispersion. The mean error in the mean, M_B in a Bernoullian Series is given by the formula:

$$\epsilon(M) = \frac{\sqrt{\epsilon^2(m_1) + \epsilon^2(m_2) + \cdots \epsilon^2(m_N)}}{N} = \frac{\sqrt{Nsp_0q_0}}{N} = \frac{\sigma}{\sqrt{N}}.$$

The mean error of the dispersion is somewhat difficult to obtain by elementary methods since it involves the determination of the mean error of the mean error. The mean error square of the mean error square may be gotten by a process similar to that of Laplace in § 65–66 by the introduction of the parameter, t, in the expression for α^3 and α^4 in $e[(\alpha - sp)^2 - spq]^2$. After several reductions this latter expression may be brought to the form: $2(sp_0q_0)^2 = 2\sigma^2$ (approx.). For the dispersion we have:

$$\epsilon(\sigma) = \frac{\sigma}{\sqrt{2N}}.$$

This formula will be proven under the discussion of frequency curves.

CHAPTER XIII.

Page 184. The Danish engineer, Andræ, discovered a similar correlation formula about the same time as Bravais.

CHAPTER XIV.

Page 196. Viewed from the standpoint of elementary errors the expression for the frequency curve, $\varphi(x)$, may be derived in the following fashion:— Let us arrange the elementary errors in small equidistant groups or intervals of magnitude, a, and assume that all the elementary errors when situated in the same interval are of equal size, an assumption which is always permissible for small values of a. This means that when r is an integral positive number all the errors located in the interval $ra - \frac{1}{2}a$ and $ra + \frac{1}{2}a$ must be of

equal size and equal to ra. The relative frequency or the probability of such errors is first of all proportional to the interval, a, and depends in the second instance upon a certain—so far unknown—function, $f(ra)$, of the quantity ra. For a particular error source, say Q_v, we may therefore express the probability of the occurrence of an error, ra, as

$$af_v(ra)$$

where f_v is the unknown function for which we make no other assumptions than those which follow immediately from the properties of mathematical probabilities, *i. e.*

$$0 \leqq f_v(ra) \leqq 1 \text{ and } \sum_{r=-\infty}^{r=\infty} f_v(ra) = 1 \text{ for } v = 1, 2, 3, \ldots s.$$

Consider now for the moment the following expression

$$F_v(\omega) = \sum_{r=-\infty}^{r=\infty} a\, f_v(ra)e^{ra\omega i} \text{ where } i = \sqrt{-1}$$

The coefficient of $e^{ra\omega i}$ in the sum is evidently the probability for the occurrence of an error ra from the error source Q_v.

The probability of the occurrence of an error ra from another error source, Q_u, may similarly be expressed as

$$F_u(\omega) = \sum_{r=-\infty}^{r=\infty} a\, f_u(ra)\, e^{ra\omega i}$$

and so on for all the s independent error sources, which we assumed to be operative on our statistical object.

The probability that the resulting sum from the various combinations into which the elementary errors from the s sources may enter is found by forming the product

$$\Phi(\omega) = \prod_{v=1}^{v=s} F_v(\omega) = F_1(\omega)\, F_2(\omega)\, F_3(\omega) \ldots F_s(\omega)$$

in accordance with the multiplication theorem of mathematical probabilities

Writing the above products as

$$\Phi(\omega) = a\,[\varphi(0) + \varphi(a)\, e^{a\omega i} + \varphi(2a)\, e^{2a\omega i} + \varphi(3a)\, e^{3a\omega i} + \ldots$$
$$+ \varphi(-a)\, e^{-a\omega i} + \varphi(-2a)\, e^{-2a\omega i} + \varphi(-3a)\, e^{-3a\omega i} + \ldots]$$

we notice that the coefficient, $a\varphi(ra)$, of $e^{ra\omega i}$ is the probability that the

elementary errors from the s error sources will enter into such a combination that their sum will fall between $r\alpha - \frac{1}{2}\alpha$ and $r\alpha + \frac{1}{2}\alpha$.

Multiplying on both sides of the above equation by $e^{-r\alpha\omega i}$ (considering $\alpha\omega$ as the independent variable) and integrating with respect to $\alpha\omega$ between the limits $-\pi$ and $+\pi$ we find that

$$\alpha\varphi(r\alpha) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \Phi(\omega)\, e^{-r\alpha\omega i} d(\alpha\omega).$$

In the above integral $\alpha\omega$ is the independent variable. If ω is chosen as the independent variable, we have

$$\alpha\varphi(r\alpha) = \frac{\alpha}{2\pi}\int_{-\pi/\alpha}^{+\pi/\alpha} \Phi(\omega)\, e^{-r\alpha\omega i}\, d\omega.$$

If now we let $r\alpha = x$ and let α converge towards zero, in which case $\alpha = dx$, we evidently find that

$$\varphi(x)\, dx = \frac{dx}{2\pi}\int_{-\infty}^{+\infty} \Phi(\omega)\, e^{-\alpha\omega i}\, d\omega$$

is the infinitely small probability that the sum of the elementary errors from the s sources will fall in the infinitely small interval $x - \frac{1}{2}dx$ and $x + \frac{1}{2}dx$.

It is evident that by introducing a new function $\psi(\omega)$, defined by the relation $\Phi(\omega) = \sqrt{2\pi}\,\psi(\omega)$, the above equation reduces to (5b) on page 195 if we let $\varphi(x) = f(x)$.

Also by writing

$$e^{\frac{\lambda_1}{\underline{|1}}\omega + \frac{\lambda_2}{\underline{|2}}\omega^2 + \frac{\lambda_3}{\underline{|3}}\omega^3 + \cdots} = \sqrt{2\pi}\,\psi(\omega)$$

we obtain the general form of the frequency curve as shown on the bottom of page 196.

Chapter XVI

Page 237. The footnote mentions McAlister as one of the earliest investigators who used the geometric mean as the most probable value of the observations. I find, however, that McAlister's work was anteceded by that of Thiele and Gram. Thiele as early as 1866 used the geometric mean as the most probable value in a series of estimates of the distances of double stars, in a Danish monograph entitled *Undersögelse af Omlöbsbevægelsen i Dobbeltstjernesystemet Gamma Virginis*. (Investigation on the movements of rotation in the double star system Gamma Vi ginis.)

Page 237. The integral in question may be evaluated as follows:—
Let $t = (z - m):n$, or $nt = z - m$ and $ndt = dz$.
Hence the integral may be written as

$$\frac{N}{n\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{rz}e^{-\frac{1}{2}\left(\frac{z-m}{n}\right)^2}e^z dz = \frac{Nn}{n\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{(r+1)(nt+m)}e^{-\frac{t^2}{2}}dt$$

$$= \frac{Ne^{(r+1)m}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{nt(r+1)}e^{-\frac{t^2}{2}}dt$$

$$= \frac{N}{\sqrt{2\pi}}e^{(r+1)m}e^{\frac{n^2}{2}(r+1)^2}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}[t-n(r+1)]^2}dt$$

If we now let $[t - n(r + 1)] = u$, we have $dt = du$, and the last expression reduces to

$$\frac{N}{\sqrt{2\pi}}e^{(r+1)m}e^{\frac{n^2}{2}(n+1)^2}\int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}}du = Ne^{(r+1)m}e^{\frac{n^2}{2}(r+1^2)},$$ since the latter integral $\int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}}du = \sqrt{2\pi}$

THE MATHEMATICAL THEORY OF PROBABILITIES
By ARNE FISHER

ERRATA

Page 292, line 2:

For $k_i = r_i k_i^2$ read $k_i = r_i k_i^1$

Page 237, lines 16 and 17 to read:

$$M_r = \int_{-\infty}^{+\infty} x^r F(x)dx = (n\sqrt{2\pi})^{-1}N\int_0^{\infty} x^r e^{-\frac{1}{2}\left(\frac{\log x - m}{n}\right)^2}dx$$

$$= (n\sqrt{2\pi})^{-1}N\int_{-\infty}^{+\infty} e^{rz}e^{-\frac{1}{2}\left(\frac{z-m}{n}\right)^2}e^z dz$$

on the assumption that z or $\log x$ is normally distributed.

Page 252, line 13: +48 to read −48.

Now in the press *Ready for distribution about August 1, 1922.*

An Elementary Treatise on Frequency Curves

And their Application to the Construction of Mortality Tables

By ARNE FISHER

English translation by E. A. VIGFUSSON

With an Introduction by Professor RAYMOND PEARL

Department of Biometry and Vital Statistics of the Johns Hopkins University

(Pp. 225+XV)

This book falls into two parts of which the first gives an elementary presentation of the theory of frequency functions along similar lines as those developed by Mr. Fisher in his book on *Probabilities.* The second part, as pointed out in Mr. Vigfusson's preface, constitutes an entirely new departure in the analysis of mortality statistics. The author has set himself the difficult task to construct a complete mortality table from mortuary records by sex, attained age at death and causes of death, but *without knowledge of the exposed to risk at various ages.* The accomplishment of this problem has been made possible by means of a *biological hypothesis and a proper classification of the causes of death upon biological principles.* Once accepted the proposed hypothesis will make it possible to study the laws of human mortality in directions which hitherto have been regarded as impossible. Mr. Fisher has applied his new method to more than 25 population or occupational groups and gives in this book the detailed results of some of his investigations in the way of 6 complete mortality tables for Michigan Males (1909-1915), Massachusetts Males (1914-1916), American Locomotive Engineers (1913-1917), American Coal Miners (1913-1917), Japanese Assured Males (1914-1917) and White

Industrial Assured Males of the Metropolitan Life Insurance Co. (1911-1916).

As a systematic treatise on frequency curves and their application to mortality studies this book should prove of great practical value not only to students of statistical methods, but to actuaries, statisticians, health officers, biologists and students of general science as well.

Comments of Specialists

**"Orthodoxy and discovery are as incompatible intellectually as oil and water are physically, a cosmic law often overlooked by our "safe and sane" scientific gentry. This book is an outstanding feature that this law is still in operation. . . . . .
It may fairly be regarded as *fundamentally* the most significant advance in actuarial theory since Halley. . . . . . It opens out wonderful possibilities of research on the laws of mortality in directions which hitherto have been wholly impossible of attack. The criterion by which the significance of a new technique in any branch of science is evaluated, is just this at the degree to which it opens up new fields of research. By this criterion Fisher's work stands in a high and secure position." (*Extract from Professor Pearl's Introduction.*)**

"Fisher's novel method has injected new blood in the old body of actuarial science." (*C. Burrau.*)

"This new and novel idea meets in reality a very frequent need. It represents a supplement to the former tools of the actuary and makes possible the utilization of a statistical material, which according to the requirements of the older systems was considered as being of no value."

(Extract from *Forsikringstidende's* report of discussion in the Norwegian Actuarial Society, June, 1920.)

"Since particularly in industrial statistics, or in general statistical inquiries under war conditions, it is easier to obtain accurate data of deaths at ages than of exposed to risk the success of the method is encouraging. . . . . The subject is one of peculiar interest at the present time."

(*Journal of the Royal Statistical Society, London, 1918.*)

From the theoretical point of view the method of Fisher is interesting. His proposal to decompose the mortality according to the different causes of death is entirely in conformity with the spirit of modern science which aims to analyze the phenomena by their differential parts. From the practical point of view the method is readily applied provided one has a double entry table of the mortuary records by age and cause of death.

(*Bulletin de l'Association des Actuaires Suisses. The Journal of the Association of Swiss Actuaries, 1919.*)